

DELHI COLLEGE OF ENGINEERING



LIBRARY
Kashmiri Gate, Delhi-110006

Accession No. _____

Class No. _____

Book No. _____

Readers are requested
to check the book and
put the signature on the
second page, if any.

Kashmiri Gate, Delhi-110006

y's delay after the due date a fine of 10 Paise per Vol. shall be charged for the first week, and 50 Paise per Vol. per day for subsequent days.

Borrower's No.	Date Due	Borrower's No.	Date Due

ELECTROMECHANICAL SYSTEM COMPONENTS

EDWARD S. CHARKEY, BME, BEE

Professional Engineer, New York

WILEY-INTERSCIENCE, a Division of John Wiley & Sons, Inc.
New York · London · Sydney · Toronto

All rights reserved. Published simultaneously in Canada.

No part of this book may be reproduced by any means, nor transmitted, nor translated into a machine language without the written permission of the publisher.

Library of Congress Catalog Card Number: 73-37364

ISBN 0-471-14700-1

Printed in the United States of America.

10 9 8 7 6 5 4 3 2

To the many aerospace engineers who were forced to make the painful transition from their chosen professions to other fields during the years

Contents

Chapter I	Analog versus Digital Systems	1
Chapter II	Sensors	32
Chapter III	Encoders	136
Chapter IV	Timers	176
Chapter V	Relays, Stepper Switches, and Stepper Motors	213
Chapter VI	Slip Rings and Potentiometers	273
Chapter VII	Motor Selection	323
Chapter VIII	Supporting Components	361
Chapter IX	System Integration	449
Appendix A	Encoder Definitions	471
Appendix B	Potentiometer Definitions	476
Appendix C	Bearing Terminology	491
Appendix D	Power Supply Terminology	502
Appendix E	Operational Amplifier Nomenclature	517
Index		525

Electromechanical System Components

Chapter I Analog Versus Digital Systems

Most sensors in use today are fundamentally analog devices; they have some form of output that is continuously representative of the quantity being monitored. Galvanometers, thermocouples, manometers, and strain gages are in this category. Digital systems employ analog sensors and convert their output to digital form at precisely controlled intervals and increments. The digital voltmeter is the most commonly used piece of digital equipment.

The principal advantages of analog systems are low initial cost, flexibility, and simplicity. Another virtue, generally not mentioned in academic circles, is familiarity. The disadvantage of this category of device is limited accuracy. It is difficult to read a dial-type indicator closer than $\pm 1\%$. This does not include the inaccuracy of the mechanism.

Digital readouts are inherently more accurate because the human reading error can be eliminated. The readouts can be made with virtually any number of significant digits. For example, a 10 V meter can be equipped with digits that correspond to tens, units, tenths, hundredths, and thousandths of a volt. The reading error would be only ± 1 unit in 10,000. It is important to note that the basic instrument error may be no better than that found in an analog instrument. However, most digital systems use techniques that produce considerably more accurate results than analog systems. The limitations of digital techniques are chiefly lack of flexibility and cost. Analog devices can be rescaled by simply adjusting a potentiometer; digital units require changing a module.

The cost element is constantly changing. At this time (1970) analog units are generally less expensive than digital units for comparable accuracies. They also have fewer parts and connections to complete. However, with further development of low-cost integrated circuit modules and increased mass production, the picture could easily be reversed in many categories within five years.

1.1. ANALOG VERSUS DIGITAL SENSORS

Here we discuss the components used in open and closed-loop servosystems, since this constitutes their principal usage. Computer components are a special type of system that is not covered in this book.

As already mentioned, most sensors are analog devices. Measurements of heat, pressure, torque, power, and accelerations are typical examples. Position sensors are uniquely suitable to analog and digital techniques. This includes angular as well as linear measurements. Some sensors monitor a process using analog techniques and convert the signal to a train of pulses that may be counted digitally. Although this puts the information in a more convenient form, the basic accuracy of the measurement is not improved. Displacement transducers are fundamentally different. If a shaft position is monitored with an encoder (see Chapter 3), shaft position changes as small as one millionth of a revolution can be detected. No analog device can compete with this level of performance. Linear encoding of machine-tool positions has also been an excellent field for digital techniques but its advantage over analog techniques is not so pronounced as that found in angular measurement. This situation can be expected to change as improved linear digital hardware is developed.

Present-day digital sensors are optical devices that sense the presence or absence of markers indicating position. They are as accurate as the marker positions they sense. These sensors are basically encoders and are discussed in more detail in Chapter 3.

1.2. AMPLIFIERS

The classical amplifier used in analog servosystems has three functions (Figure 1a):

1. *Error detector.* One section of the amplifier module is used as a summing network to establish the difference between the command and feedback signals.

2. *Error amplification.* The second section is used to amplify the error signal so that it can drive some corrective electromechanical device. The resolution of the error detector-amplifier sections determines the response of the servosystem.

3. *Impedance matching.* The output impedance of the amplifier module must be designed so that the impedance of the servomotor or other corrective element does not load the amplifier under normal dynamic conditions.

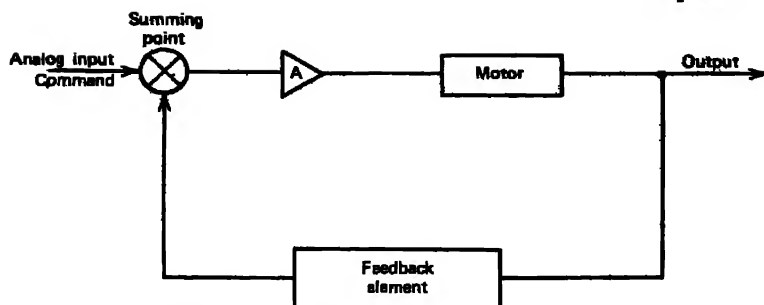


Figure 1a. Typical analog servosystem.

Digital servosystem amplifiers differ from analog units in several important aspects (Figure 1b). The error detector section is equipped with supplementary logic circuitry to cause the error input to the amplifier to occur at one particular error value. For example, an analog error detector may function continuously for error signals between 1 mV and 1 V; the digital system may be designed to function at 17 mV or some other discrete value. Below this value the amplifier is inoperative; above this value it drives the servomotor at some finite power level. The input to this type of system must be in digital code, and the feedback signal is derived from an encoder. The output of the amplifier may be fundamentally an "on-off" operation or a conventional linear output that is proportional to the number of digital increments in excess of an established value. For example, suppose that the digital error detector is designed to function when the difference between the command signal and the error signal is 27 digital units. The amplifier can be designed so that it produces a linear output of 0.5 V per digit above this value.

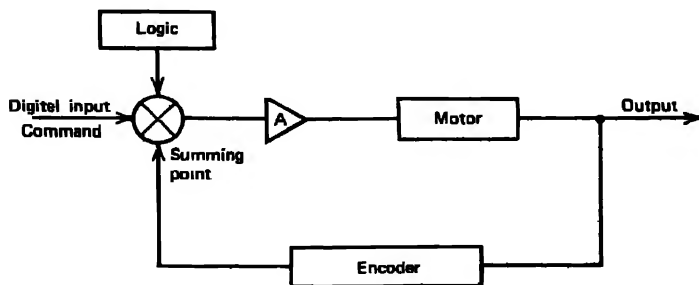


Figure 1b. Digital servosystem.

4 Analog Versus Digital Systems

Currently available hardware can be used for innumerable special functions such as the one illustrated. Virtually any of the logic used in Fortran digital programming can be incorporated in the design. The unit can be designed to function below or above a given digital value, or across a given band of values. It can also be programmed to perform at positive or negative values. When the amplifier is used in conjunction with a stepper motor, a fixed relationship can be established between the error signal and the output of motor. This is frequently employed in digital displays.

1.3. OUTPUT DISPLAYS

Analog output displays have remained substantially unchanged for years. The sensitivities have improved, the bearings have been refined, and the numerals have become easier to read and more attractive, but fundamentally they still resemble a clock face. During the past 10 years, because of stimulus from the human factors groups, vertical-scale analog displays have come into general use. The face is rectangular, and the printer or scale moves up and down rather than circularly. Digital displays are not much more than 20 years old and, by contrast, have continually been in a state of evolution. They have changed from the form used in "EPUT" meters to the circular nixie tube* and then to in-line designs where various combinations of numerals can be formed by selecting the proper variety of filaments or light-emitting diodes. The salient feature of digital readout displays is that they completely eliminate reading error; the inherent accuracy built into the system is available at the output. In contrast, the reading error of analog systems ranges from 1 to 5%. The principal disadvantage of digital displays is cost. At the present time they are not priced competitively with analog devices.

To utilize the best features of analog or digital systems, it is frequently necessary to convert from one system to the other. The resultant hybrid exhibits some advantage in price, accuracy, or size that is unattainable with a conventional approach. The two possibilities are analog-to-digital and digital-to-analog conversions.

1.4. ANALOG-TO-DIGITAL CONVERSION TECHNIQUES

The most common conversion used in modern instrumentation is analog-to-digital; this provides the benefits of digital readouts that we discussed earlier.

* The term nixie tube is used in the generic sense, referring to a multifilament tube-type device that displays numbers or letters. The term NIXIE TUBE is a registered trademark of the Burroughs Corporation.

The fundamental problem associated with the conversion of analog signals to discrete digital approximations is fidelity. How are a series of complex waveforms reduced to reasonably accurate straight-line approximations? This process is called quantizing and is best described as the method of converting an analog to a digital signal. Quantizing consists of two processes:

1. Digitally matching the amplitude of a series of discrete points on the analog function.
2. Matching the frequency of the analog function by sampling the amplitude at a proper rate.

If points 1 and 2 are performed well, the fidelity of the quantizing process is acceptable.

To illustrate the amplitude matching process, assume the analog waveform to be a sine wave. Consider the amplitude at one point that is stopped in terms of time (Figure 2). Assume that the amplitude at this point is $10\frac{1}{2}$ units. The digital units available for this value are in whole units; therefore the least significant bit (LSB) is 1 unit. The closest approximation possible is 10 or 11 units and the associated error is $\pm\frac{1}{2}$ bit. If the LSB were $\frac{1}{2}$ unit, the approximation would be perfect at this point. Practically, the numbers never occur in such convenient form. Consequently, the smaller

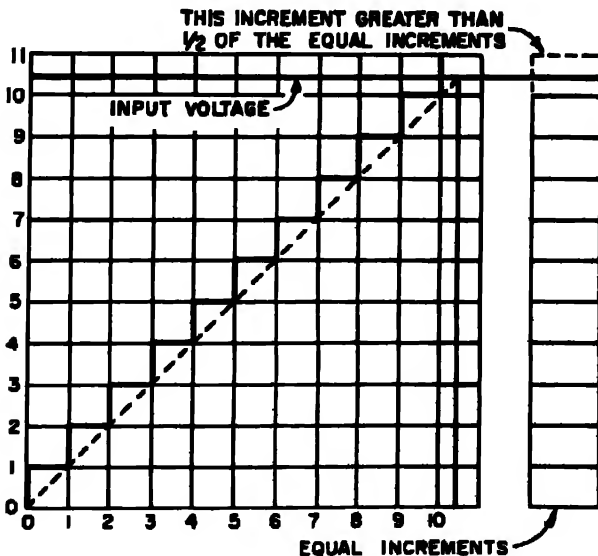


Figure 2. Quantized voltage and best fit of equal increments. (From Reference 1, p. 302.)

the LSB the better the approximation, since the basic maximum error is $\pm \frac{1}{2}$ LSB.

The rate at which discrete points along the sine wave are analyzed is called the aperture rate, and is expressed in conversions per second. The higher the aperture rate, the better the fidelity. This is similar to the classical mathematical illustration of integration where the total integral consists of the sum of a number of slices whose width is expressed as dx . When determining how fast the aperture rate should be, a good rule of thumb is that the rate must be at least twice the frequency of the highest harmonic in the waveform. Sometimes this determination is a difficult problem, since many analog waveforms are rather complex. A reasonable approach is to analyze typical data with a harmonic wave analyzer and determine how many harmonics are essentials for the task being performed. The error due to aperture rate may be determined as follows.

Assume the analog voltage $V = A \sin \omega t$

$$\frac{dV}{dt} = \omega A \cos \omega t$$

When $\omega t = 0$, $dV/dt = \omega A$.

This is the point at which the curve changes most rapidly. Let $A = 1$ V (peak to peak) and $\omega = 2\pi f$; then $dV/dt = 2\pi f$. If we replace dV with ΔV and dt with Δt (the aperture time), we have

$$\Delta V = 2\pi f \Delta t$$

The error, in voltage, due to aperture time is

$$\% \text{ error due to } \Delta V = 2\pi f \Delta t \times 100$$

For example, assume that the A/D converter available has an aperture time of $10 \mu\text{sec}$ and the highest harmonic present in the waveform is 500 Hz. The aperture rate should be $2 \times 500 = 1000$ Hz. The resultant error is

$$\text{error} = 2\pi \times 10^3 \times 10 \times 10^{-6} \times 10^2 = 6.28\%$$

Example: find the aperture time required to digitize a 1000 Hz signal with 0.1% accuracy.

$$0.1 = 2\pi \times 10^3 \Delta t \times 10^2$$

$$\Delta t = 0.159 \times 10^{-6} \text{ sec}$$

In summary, for a good digital approximation of an analog signal, both the quantizing voltage and aperture rate must be carefully evaluated.

1.4.1. Types of A/D Converters

Numerous methods have been proposed for categorizing A/D converters. One method is to distinguish between programmed and nonprogrammed ones. In programmed units the conversion is performed in a given number of steps, with each step clocked to last a fixed time interval. The non-programmed type may require a sequence of events to occur before the conversion is complete. This sequence is not in fixed time steps and depends on the response time of the conversion circuitry.

Another classification is into feedback and open-loop units. Open-loop units compare the analog input voltage and a reference voltage. The result is a digital output equivalent to the difference. Closed-loop converters generate an internal analog voltage that is proportional to the applied analog voltage. The digital output is proportional to the internal analog voltage. The external and internal voltages are compared; if they are unequal, the internal voltage is increased until a null condition is reached. As this process proceeds, the digital output is increased progressively.

A third method of classification is by the number of conversions per second. This determines the number of LSBs* that can be employed in the amplitude matching process. The classification is as follows:

Low-speed: up to 10 conversions/sec.

Medium-speed: 10 to 1000 conversions/sec.

High-speed: 1000 to 100,000 conversions/sec.

Video-speed: 100,000 to 10,000,000 conversions/sec.

The low-speed categories use stepper switches and reed relays. All the others use solid-state circuitry.

The last category groups the converters by two types:

1. Those using capacitor-charging circuits that digitally encode the time required to charge the capacitor to the required level.

2. Those employing a discrete voltage comparison circuit that depends on the generation of voltages whose amplitudes are equivalent to digital values. It is basically a closed-loop device that compares the voltage generated with the input analog voltage and produces an output digital signal that is proportional to the difference.

The last categorization is used in this book.

* The term LSB (least significant bit) is discussed in detail in Chapter 3.

1.4.2. Capacitor-Charging A/D Converters

Three basic types of capacitor-charging circuits will be discussed. There are numerous commercial variations of these designs, each with a significant advantage, but the circuits discussed below are the common starting point.

1.4.2.1. Voltage-to-Frequency A/D Converter. (Figure 3) The analog input voltage is converted to a proportional constant current, which is integrated by amplifier K and its associated "RC" network. (The resistor is part of the current converter network.) The integration process produces a series of ramps. The peak value is controlled by two analog comparators; when the ramp voltage generated by amplifier K equals the reference voltage on the comparator amplifiers, one of them generates a pulse that resets the integrator input to zero. The number of ramps generated per second, or frequency, is therefore proportional to the analog input and the reference voltages $+V_R$ and $-V_R$. The number of pulses generated for a given time interval is totaled by a binary counter. The system is scaled so that a given number of pulses per unit of time are digitally equal to a prescribed scale voltage.

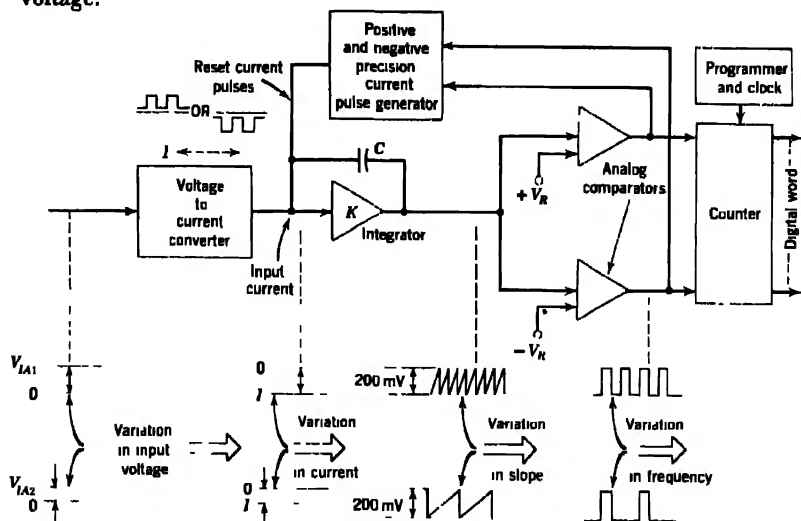


Figure 3. Voltage-to-frequency conversion. (From Reference 2.)

1.4.2.2. Pulse Width Modulator A/D Converter. (Figure 4) Pulse width modulator designs are used when low conversion rates, medium accuracy, and relatively few components are desired. When switch S_1 is moved to the reset position, the constant current source I starts to linearly

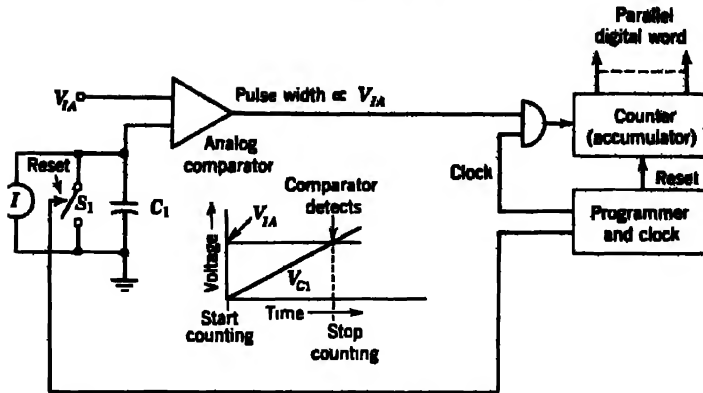


Figure 4. Pulse-width modulator A/D converter. (From Reference 2.)

charge capacitor C_1 . The reset operation also starts the programmer and clock. The capacitor voltage is directed to one input of an analog comparator. The other input to the comparator is the analog input, V_{IA} , that is being digitized. So long as there is a difference between these two signals, the comparator generates an output pulse. The width of the pulse is a function of the time required to equalize the two inputs to the comparator—hence the name pulse width modulator. The pulse generated and the clock signals are directed to a NAND gate. While both inputs are present at the gate terminals, the counter will accumulate digits. When the output from the comparator drops to zero, the NAND gate stops passing clock pulses and the counter is frozen until it receives a reset pulse. The final count in the accumulator is the digital equivalent of the input analog waveform at one discrete point on the analog waveform.

The overall accuracy of this circuit is about $\frac{1}{2}$ to 1%. It is relatively slow compared to other devices. For example, for a 7-bit conversion,

$$[1 \ 2 \ 4 \ 8 \ 16 \ 32 \ 64]$$

the counter requires 127 pulses before reaching a full-scale count. If a clock frequency of 100 kHz is used, this would require 127×10^{-5} or 1.27×10^{-3} sec. One advantage of this system is that when the pulse width generated is sent from one piece of equipment to another, only the time position of the leading and trailing edges of pulses are important; variations in pulse amplitude and pickup are not critical. The largest errors inherent in this type of circuitry are variations in the capacitor, such as temperature stability and changes in value with aging. The constant-current source is also not

perfect and limits the linearity of the voltage buildup across the capacitor. Figure 5 diagrammatically shows the events in the circuit operation.

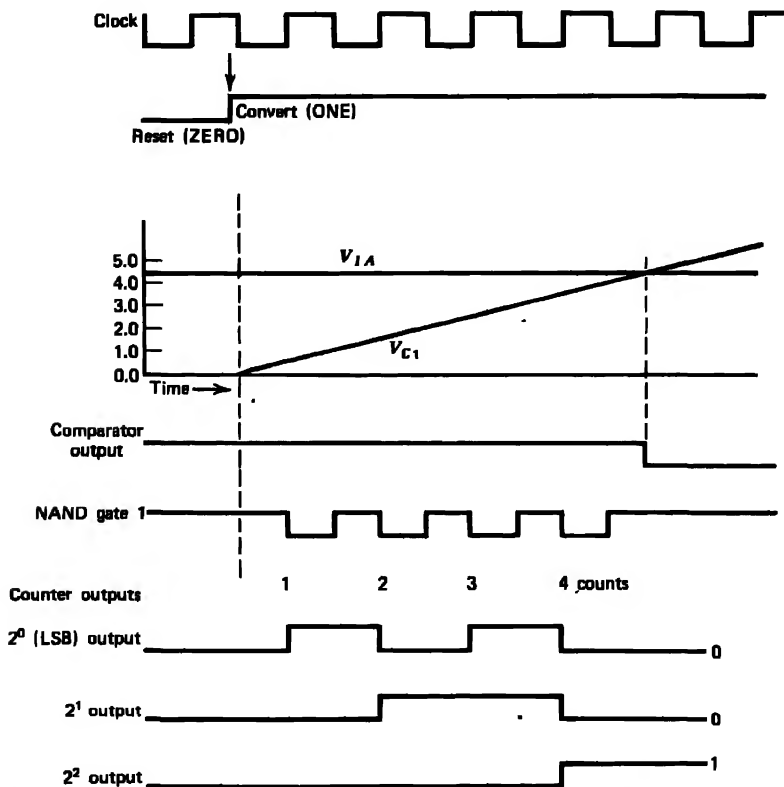
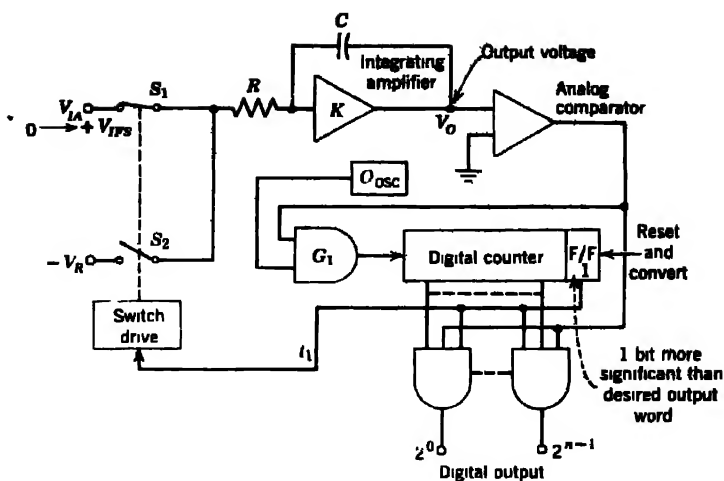
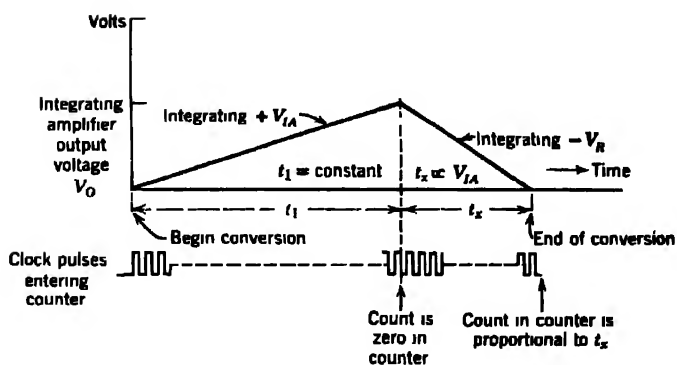


Figure 5. A/D conversion timing diagram for a 3-bit output (Pulse-width-type A/D converter). (From Reference 2.)

1.4.2.3. Up-Down Integrator A/D Converter. This technique is more accurate than other capacitor-charging converters because it cancels out the errors associated with the capacitor. It provides relatively low conversion rates and is somewhat more complex than the other capacitor circuits (Figure 6). The basic method of operation consists of integrating the input analog voltage for a fixed period, then integrating a negative reference voltage and measuring the time required to reach zero voltage again. The



(a)



(b)

Figure 6. (a) Block diagram of up-down integrator A/D converter; (b) timing diagram. (From Reference 2.)

second period is proportional to the amplitude of the original analog input. Since the same integrator circuit is used for both integrations, any variations in capacitor parameters cancel out.

The detailed process is as follows: Upon receipt of a reset signal, flip-flop 1 is set in its zero position, as are all elements of the digital counter. This causes the switch drive element to close switch S_1 and open S_2 . The input analog voltage, V_{IA} , is then supplied to the integrating amplifier. The instant the output of amplifier K changes from zero, the analog comparator amplifier supplies a voltage to NAND gate G_1 , enabling it to conduct pulses provided by the reference oscillator. The digital counter will then count pulses so long as the integration process produces an output voltage. The process continues until all the digits in the counter read one [111...111]. The next oscillator pulse causes the counter to change to all zeros, which triggers the flip-flop to its "1" state. The output of the flip-flop causes the switch drive to change its state, so that S_1 opens and S_2 closes. This applies a negative reference voltage, $-V_R$, to the integrator circuit and causes the output of the integrator to diminish until it reaches zero. At the zero voltage level the output from the analog comparator becomes zero and NAND gate G_1 will no longer pass the oscillator pulses. The number in the digital counter is the analog of the time required to go from peak amplitude to zero voltage, which is a function of the amplitude of the input analog voltage magnitude. Therefore, the greater the amplitude of V_{IA} , the longer the period defined in Figure 6 as t_x . The process is independent of the integrator time constant and its variations. It is essentially a function of the oscillator stability. The principal errors are the comparator and analog switching errors which are quite low. The requirement for 2 integrations per input analog voltage is responsible for the low conversion rate of this technique.

1.4.3. Discrete Voltage Comparison A/D Converters

The fundamental difference between integrator and discrete converters is that the former generates a linear voltage that is compared to the analog voltage while the latter generates a series of discrete voltage steps.

1.4.3.1. Counter Ramp A/D Converter. This device produces high accuracy at low conversion rates with low circuit complexity (Figure 7). A reset pulse directed to the counter sets it to zero and clears the output of the D/A converter; consequently, there is an immediate output from the analog comparator. This voltage applied to NAND gate G_1 permits it to pass clock pulses to the counter. For each count that the D/A decoder receives from the counter its output increases by a uniform amount. The output waveform that is generated resembles a staircase until it is slightly greater than the input analog voltage; at this point the output of the comparator changes state and inhibits the NAND gate from passing any additional

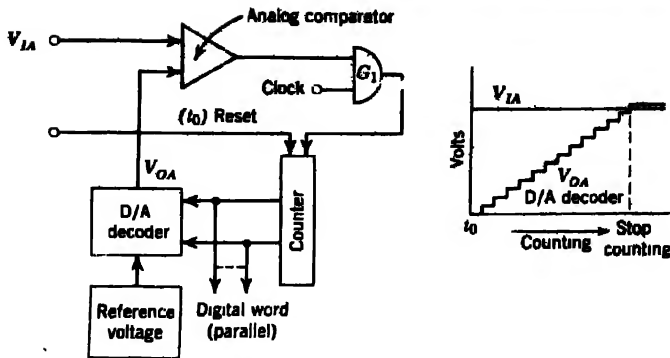


Figure 7. Counter ramp A/D converter. (From Reference 2.)

clock pulses. The reading on the counter is then the digital equivalent of the input voltage. If the counter has n runs of digits, the circuit must generate $2^n - 1$ clock pulses for full-scale readings. This feature makes this technique relatively slow and unsuitable for high-speed applications.

1.4.3.2. Successive Approximation A/D Converter. This converter is used where medium conversion rates, high accuracy and medium complexity are desirable. The basic scheme consists of generating a series of step voltages starting with the most significant bit (MSB),* comparing it with the input analog voltage and then proceeding with the next most significant bit (NMSB) until all the bits have been accounted for. By most significant bit we mean the bit with the greatest numerical value. For example, in the binary number 1001, the one on the left represents 8, the first zero represents zero 4's, the next zero represents zero 2's, and the last figure represents unity for a total value of 9. The leftmost one is then the MSB, the first zero is the NMSB, and so on (Figure 8). A more detailed discussion of coding appears in Chapter 3. The operation is initiated when a reset pulse sets the output of the D/A decoder, V_{OA} , to zero and the input analog voltage, V_{IA} , is applied. The analog comparator generates an output that is applied to the programmer and NAND gate G_1 . The programmer then initiates a pulse equivalent to the MSB. The D/A converter then generates a proportional voltage. If it is less than V_{IA} , an output from the comparator is applied to the NAND gate and is transmitted as the MSB to external recording circuitry. If the MSB applied to the comparator is greater than the input analog voltage, the comparator changes state and the output from the NAND gate is a logical zero. The programmer then shifts to the NMSB, which, together with

* The term MSB (most significant bit) is discussed in Chapter 3.

14 Analog Versus Digital Systems

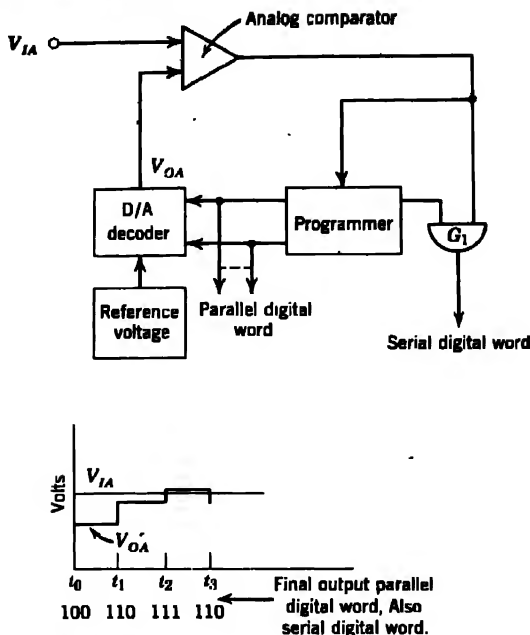


Figure 8. Successive approximation A/D converter. (From Reference 2.)

the MSB, is compared in turn with the input analog voltage. The process is repeated for each significant digit in the code until the numeral is complete. This process is much faster than the counter ramp technique, since only one operation is required for each significant digit.

The terms serial digital word and parallel digital word found in Figure 8 are defined as follows. A serial digital word or number can be transmitted 1 significant digit at a time, starting with the MSB. Transmission of the data can be started while the A/D conversion is still in process. Parallel digital data require that all digits be read simultaneously and cannot be read until the A/D conversion is completed.

Accuracies of better than $\pm 0.005\%$ are possible with successive-approximation A/D converters. A detailed study of circuitry can be found in Reference 2. Although the circuitry is considerably more complex than that of the ramp technique, integrated circuit methods have made it economically sound.

1.4.4. Other Types of A/D Converters

So far we have discussed the two generic types of A/D converters—capacitor-charging units and discrete-voltage types. There are numerous

variations on these basic designs, each with a special advantage in accuracy, price, or technique. Very often a design is based on some type of proprietary technology developed by a company for an allied product. Four additional types of converters are now presented to round out the description of available hardware:

1. Operational amplifier A/D converters.
2. Emitter-follower A/D converters.
3. Parallel A/D converters.
4. Parallel/serial A/D converters.

This is not a complete list, since new designs are constantly being developed based on evolving solid-state technology.

1.4.4.1. Operational Amplifier A/D Converter. This converter is a good compromise when medium conversion rate, accuracy, and complexity are desired (Figure 9). The input analog voltage, V_{IA} , is directed to comparator C_1 where it is compared against one half of the full-scale voltage, V_{IFS} . (The full-scale voltage is the amplitude required to set the output of the counter to its full-scale reading. For example, a 4-bit binary counter would read 1111 at full scale.) If the input voltage is greater than $\frac{1}{2}V_{IFS}$, comparator C_1 produces a digital "1" output that is recorded by external circuitry as the MSB. The output is also used to open switch S_1 which permits $-\frac{1}{2}V_{IFS}$ to be applied to the terminal of operational amplifier K_1 . This value is subtracted from the input V_{IA} and the difference is inverted by amplifier K_1 and applied to the input of the second comparator C_2 . It is now compared with $-\frac{1}{4}V_{IFS}$, the next most significant bit. If the input is greater than the reference voltage, C_2 produces a logical "1" output; if it is less, C_2 produces a logical zero output. For all bits, if the comparators do not detect it, $|V_{IA}| > |(1/2^n)V_{IFS}|$, the input voltage is passed through the operational amplifier with a gain of 1, since switches S_1, S_2, \dots are not opened. This process is continued down to the LSB circuit. Practical circuitry uses FET components as the switches.

1.4.4.2. Emitter-Follower A/D Converter. Emitter-follower circuits are used to replace the operational amplifiers of the previous design. The advantages are very high conversion rates and low cost, but lower accuracies also appear (Figures 10a and 10b). The output of each emitter-follower contains a series resistor connected to the input of the next emitter-follower. A current source is connected to the succeeding emitter-follower input. The application of a digital "1" voltage to the input turns on the current source connected to the input of the succeeding bit's emitter-follower. The current causes a fixed drop in voltage across the resistor R_1 . This voltage drop subtracts the voltage weight of the previous bit from the analog input voltage to the emitter-follower for the next less significant bit. The next bit's comparator then compares its reference against the difference analog signal.

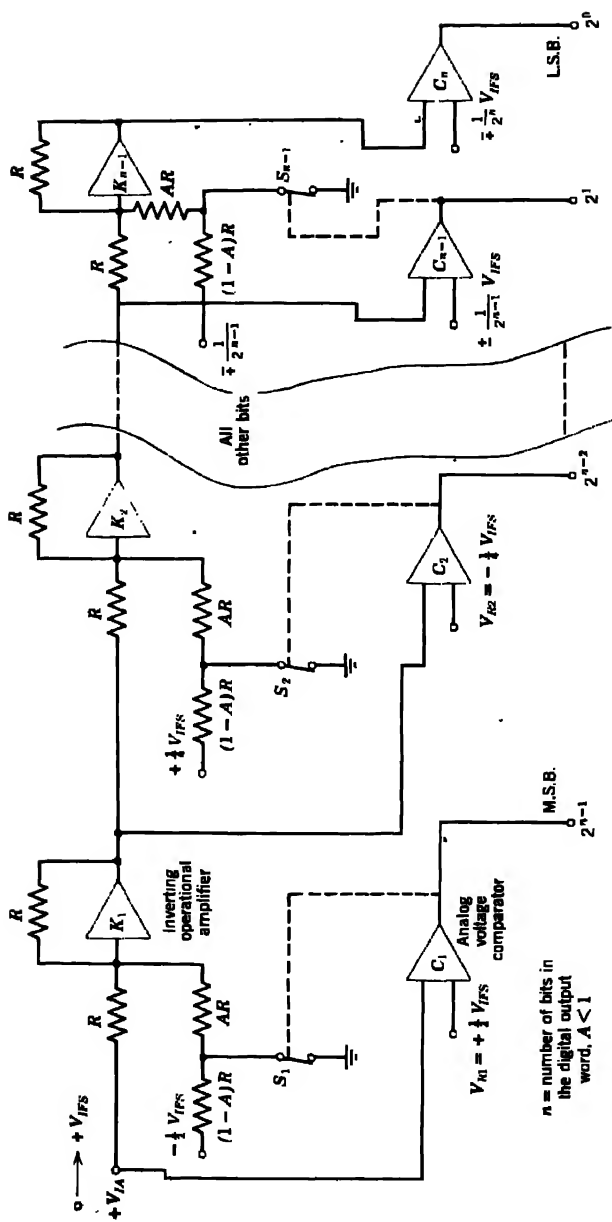
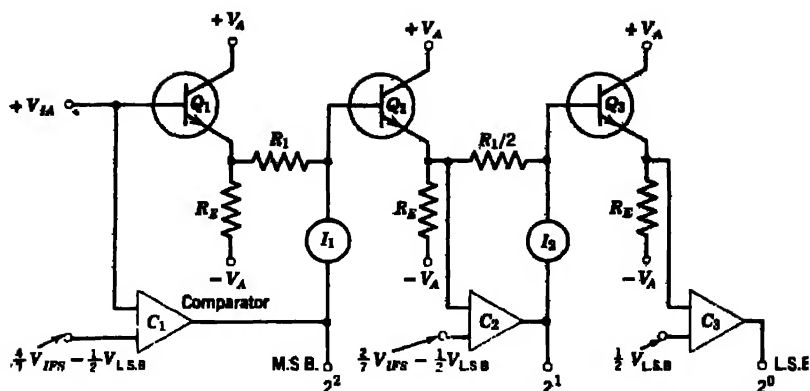


Figure 9. Basic operational amplifier A/D converter. (From Reference 2.)



where

$$V_{LSB} = \frac{1}{2^{n+1}} V_{IFS}$$

Figure 10a. Emitter-follower direct comparison A/D converter. Note that I_1 is "on" for a digital "1" from the digital comparator; $I_1 = I_2$. Also for this 3-bit A/D converter, $I_1 R_1 = \frac{1}{2} V_{IFS}$; $I_2 R_1/2 = \frac{1}{2} V_{IFS}$. (From Reference 2.)

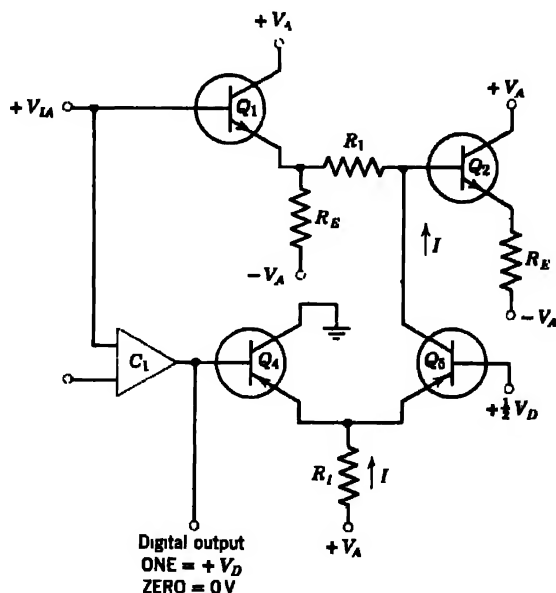


Figure 10b. Possible constant-current source circuit for emitter-follower direct comparison A/D converter. $I = (+V_s - V_{BES} - V_D)/R_1$. (From Reference 2.)

Some of the problems associated with this circuit are variations in the base-to-emitter voltage, temperature and current variations in the coupling resistors, and parasitic capacitive effects.

1.4.4.3. Parallel A/D Converter or Simultaneous Converters.

This type provides the highest speed of any converter. The devices use an analog comparator for every quantization level in the encoded digital word. Since the conversion is performed in one step, rates of hundreds of megabits per second can be achieved. However, the amount of equipment required is almost doubled for each additional binary bit of resolution. Consequently, this technique is used where high speed and low resolution are required, typically in 3- to 6-bit conversion systems. A typical 3-bit system is shown in Figure 11. A 3-bit system requires 7 comparators, each biased with a given

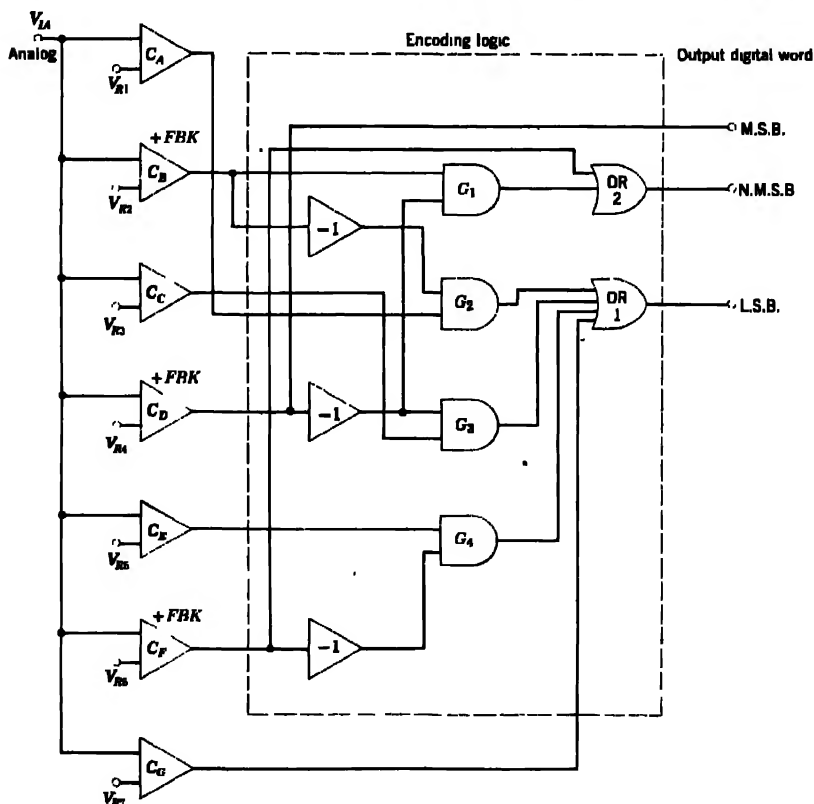


Figure 11. Parallel 3-bit A/D converter. The +FBK indicates comparators that should be regenerative to prevent ambiguities. Note that their change causes a change in more than 1 bit at a time. (From Reference 2.)

reference voltage. When the input analog voltage is applied, it is simultaneously connected to each of the comparators. The resulting "1" or "0" outputs are grouped logically according to equations developed through the use of Boolean algebra. For example, the equation governing the output from the LSB is as follows:

$$A \cdot \bar{B} + C \cdot \bar{D} + E \cdot \bar{F} + G \rightarrow \text{LSB}$$

The logical terminology is defined in the following way:

$+$ = OR

\cdot = AND

\bar{A} = output of C_A at "0" level

A = output of C_A at "1" level

The relationship between the digital output and the comparison levels is shown in Figure 12.

Ambiguities can occur in parallel A/D converters around the threshold values for some of the comparators. One way of preventing this is to make some of the comparators regenerative. This means positive feedback is

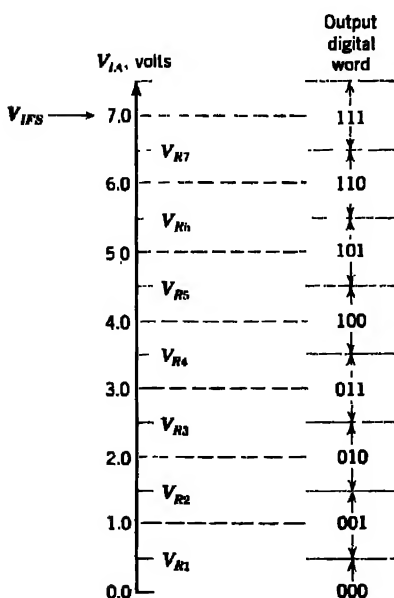


Figure 12. Relationship between V_{IA} and the digital word output. (From Reference 2.)

designed into the circuit so that when a comparator begins to change its state the positive feedback carries it completely through its transition to opposite state. Then, with the comparators always being definitely in "1" or "0" state except for the small transition time, the decoding logic will not encode ambiguously. This problem is discussed in greater detail in Chapter 3. In Figure 11 the amplifiers marked +FBK are feedback units.

1.4.4.4. Parallel/Serial A/D Converters. These converters are a compromise to obtain some of the resolution of serial converters and the speed of parallel units (Figure 13). The process consists of an initial parallel conversion to determine and store the MSBs and a succeeding conversion of circuitry to determine the value of the LSBs.

The convert command signal sets F/F 4 in the "1" state, turning I_C (the coarse constant current source) and presetting F/F's 1, 2, and 3 to "0" state. The preset signal from F/F 4 is DC-coupled to the flip-flop inputs, holding them in the "0" state as long as the signal is present. The output of the D/A decoder is then at zero volts. The seven reference voltages are set up for the conversion of the three MSBs by the voltages developed across resistors R_1 through R_6 by I_C . At the end of the delay time, t_1 , F/F 4 is reset to the "0" state, conditioning gates 1, 2, and 3 so that if a "1" exists on an encoding logic output line, a triggering signal is coupled through the gate to set the driven flip-flops. The digital value for the three MSBs is thus stored in F/F 1, 2, and 3. The process for resolving the analog voltage down to the three LSBs is as follows. The 3-bit D/A converter generates an analog voltage proportional to the value of the three MSBs, thereby raising the voltage of the resistor divider (R_1 to R_6) to that level. Additionally, F/F 4 turns I_C off and I_F on. The reference voltage at the resistor-divider terminals then represents the seven possible discrete levels (for the three LSBs) between the value of the previously encoded three MSB partial word and the next highest three MSB words. The comparators detect the reference voltages exceeded by the incoming analog voltage. The comparator outputs are encoded, as before, to 3 parallel digital bits of information. Gates 4, 5, and 6 are enabled by F/F 4 to pass the 3-bit information to the output. The 6-bit digital word can then be read out, and the process repeated by the application of another convert-command signal.

1.5. DIGITAL-TO-ANALOG CONVERTERS

The process of converting from digital to analog format is basically simpler than the inverse operations just discussed. This is somewhat like a discussion of differential and integral calculus; the processes are inverses of each other but the methods of achieving the results are markedly dissimilar.

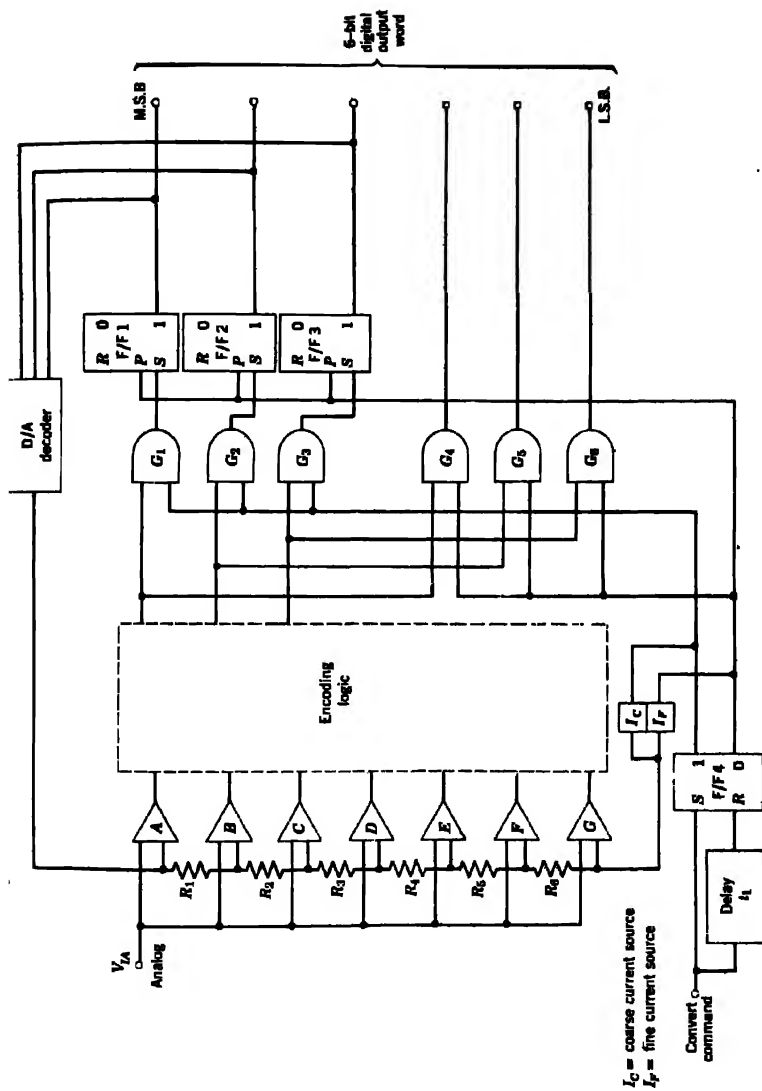


Figure 13. Parallel/serial A/D converter (6-bit). (From Reference 2.)

A D/A converter, in its simplest form, consists of a series of switches and a voltage divider network that provide analog output signals proportional to the switch activated. Figure 14 illustrates the operation of a 3-bit unit. To provide for proper polarity, 4 digits are used.

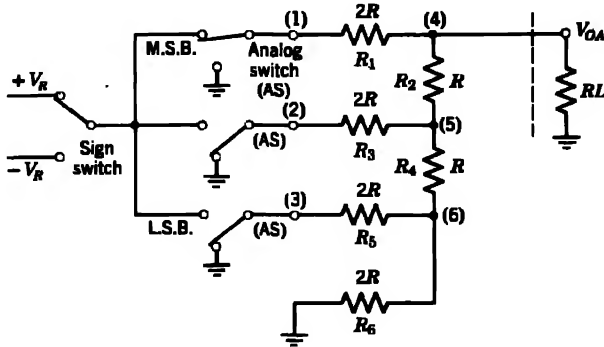


Figure 14. $2R,R$ resistor ladder D/A decoder (digital word 1100). (From Reference 2.)

The first digit indicates polarity; 1 signifies a positive number and zero indicates a negative number. The second digit represents the state of the MSB; the third, the NMSB; and the fourth, the LSB. Thus the binary digital word 1100 means a positive value of 4 (to the base 10). A voltage delivered at junction point 4 will be attenuated the least with respect to the load, and a voltage introduced at point 6 will be attenuated the most. It is therefore possible to supply identical voltages to a properly designed divider network and to obtain amplitudes that are proportional to the numerical value of the switches supplying the input. In this discussion it is assumed that the load resistance, R_L , is made much greater than the impedance of the network to prevent loading errors.

A typical D/A converter is composed of four major elements:

1. *Logic circuitry* consists of a series of NAND and NOR gates plus flip-flops that receive various signals from the digital input device and sort them into categories that will uniquely activate one or more output switches.
2. *Resistor network* converts equal voltages, supplied at various terminals, to voltage amplitudes proportional to their intended values.
3. *Analog switches* are composed of a series of high-speed solid-state circuits that open and close circuits according to the dictates of the logic circuitry. The sign switch is also considered part of this category.
4. *Reference voltage* consists of positive and negative voltages, designated as $\pm V_R$ in Figure 14, that supply the polarity indication for the circuit.

The technique described above is sometimes referred to as a ladder-type D/A decoder. The resistance at each node in the network, looking away from the load, is equal to R (Figure 14). At node 6 the impedance is equal to the parallel combination of R_5 and R_6 which equals R . At node 5 the resistance is equal to R_5 in parallel with R_4 plus node 6; again the value is R . Selection of the value of the resistors used is dependent on the magnitude of R_L ; the load must always be much higher than network resistance. Another reason for maintaining as low a network resistance as possible is to keep the response time down. The network and the stray capacitive effects in each element tend to limit the conversion speed.

The general expression for the output voltage available from a ladder-type converter of n bit length is as follows:

$$V_{OA} = \left[\frac{1}{2} D_1 + \frac{1}{4} D_2 + \frac{1}{8} D_3 + \frac{1}{16} D_4 + \cdots + \frac{1}{2^n} D_n \right] \frac{V_R R_L}{R + R_L}$$

where V_{OA} = the output voltage across the load

D = the state of the digital input of a particular bit. If the digital input for a particular bit is "1", $D = 1$; if the digital input for a particular bit is "0", $D = 0$

V_R = reference voltage

R_L = load resistance

R = resistance of the network

One of the advantages of using this type of decoder is that all the resistors in the network are either R or $2R$ in magnitude. It is therefore simple to obtain resistors with closely matched temperature coefficients, since their numerical values differ by only a factor of 2.

Weighted resistor D/A decoders offer another approach (Figure 15). The magnitude of the resistor is inversely proportional to the value of the digital bit decoded. The R accommodates the MSB, and $4R$ handles the LSB.

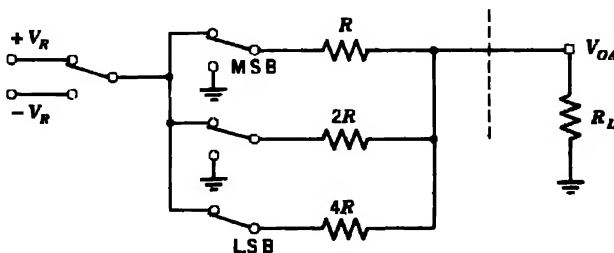


Figure 15. Three-bit plus sign weighted-resistor D/A decoder. (From Reference 2.)

When the switch in each branch is closed, the V_R signal is transmitted across the appropriate resistor and develops a logical 1 signal; when the switch is grounded the signal is zero. The circuit shown is binary +111, or 7 to the base 10. The equivalent analog value for digital words of n -bit length is as follows:

$$V_{OA} = \left[\left(\frac{2^{n-1}}{2^n - 1} \right) D_1 + \frac{1}{2} \left(\frac{2^{n-1}}{2^n - 1} \right) D_2 + \frac{1}{4} \left(\frac{2^{n-1}}{2^n - 1} \right) D_3 + \dots \right. \\ \left. + \left(\frac{1}{2^n - 1} \right) D_n \right] \left[\frac{V_R R_L}{R_O + R_L} \right]$$

where $R_O = [2^{n-1}/(2^n - 1)]R$. The other symbols used are the same as those for the ladder network. The first term represents the MSB and the last term the LSB. Sometimes the network is equipped with an extra resistor connected in parallel with the LSB with one end grounded; its magnitude is equal to the LSB resistor. This makes the output impedance equal to $R/2$ no matter how many weighted resistors are connected in parallel. This can be checked by calculating the impedance at each node (Figure 16).

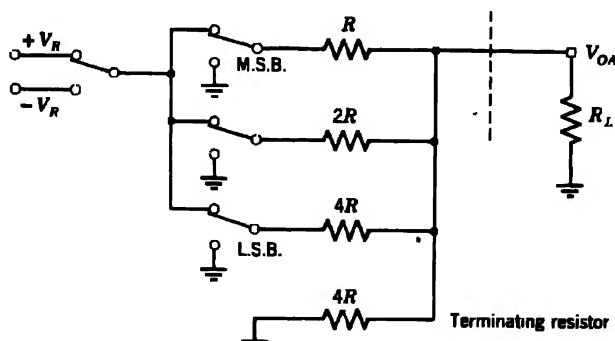


Figure 16. Weighted-resistor D/A decoder with terminating resistor. (From Reference 2.)

The principal advantage of the weighted resistor technique is that it is possible to use lower-rated and less expensive components in the less significant bit branches. Only the MSB and those branches that are reasonably close to it require components with substantial ratings. When a system consists of 10 or more bits, the savings can be impressive. This includes the resistors as well as the semiconductors that comprise the switches. However, since the

value of the resistors used in the networks vary substantially, temperature compensation is usually more of a problem than with the ladder design.

Weighted-current D/A decoders are used when very high conversion speeds coupled with medium accuracy are required (Figure 17). Each bit has a

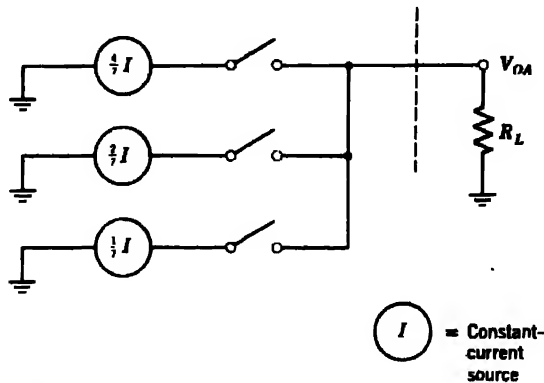


Figure 17. Weighted-current D/A decoder. (From Reference 2.)

branch equipped with a constant-current source that develops a proportional voltage across the load resistor R_L . A basic limitation of this technique is that when a base output voltage is required it is difficult to attain a low output impedance from the constant-current sources. Figure 18 shows the output voltage obtainable for various digital word inputs.

The *single-valued-current D/A decoder* combines a constant-current source and a modified ladder network. This method uses a single-value constant-current

Digital Word			Analog Output, V_{OA}
M.S.B.		L.S.B.	
0	0	1	$\frac{1}{7} IR_L$
0	1	0	$\frac{2}{7} IR_L$
1	0	0	$\frac{4}{7} IR_L$
1	1	1	IR_L

Figure 18. Weighted-current D/A decoder—digital words versus analog output voltage. (From Reference 2.)

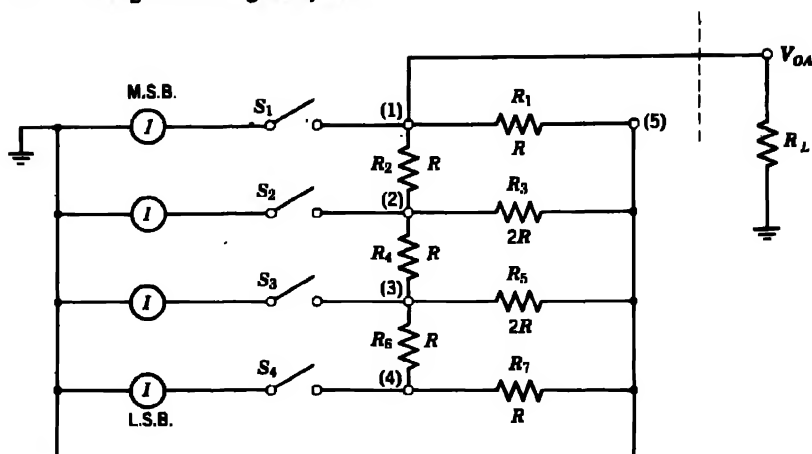


Figure 19. Single-valued-current D/A decoder. (From Reference 2.)

source that simplifies the circuit design (Figure 19). The ladder design is composed of R and $2R$ resistors that are easily matched. When the MSB switch, S_1 , is closed, one third of the current passes through R_1 ; when S_2 is closed, one sixth of the current passes through R_1 ; and so on. The general expression for the output voltage of an n -bit circuit is as follows:

$$V_{OA} = \left(\frac{2}{3} D_1 + \frac{1}{3} D_2 + \frac{1}{6} D_3 + \cdots + \frac{1}{2^{n-1}} D_n \right) \frac{IRR_L}{\frac{2}{3}R + R_L}$$

The symbols are the same as for previous equations.

A technique that is commonly used for efficiently adding the outputs of a D/A network is the null summing method (Figure 20). This uses an operational amplifier to produce the null. The mode of operation is as follows. If we assume that the amplifier has nearly infinite gain (10,000 to 100,000) and its input impedance is very high, the feedback current, I_f , is equal to the sum of the individual branch currents I_1 , I_2 , I_3 , and so on. This assumes that no current is diverted to the amplifier.

$$I_f = I_1 D_1 + I_2 D_2 + I_3 D_3 + \cdots + I_n D_n$$

$$I_f = \frac{V_1 D_1}{R} + \frac{V_2 D_2}{2R} + \frac{V_3 D_3}{4R} + \cdots + \frac{V_n D_n}{2^{n-1}R}$$

This is based on the weighted resistor design shown in Figure 20 where D_1 , D_2 , and so on, represent the states of the switches (1 or zero).

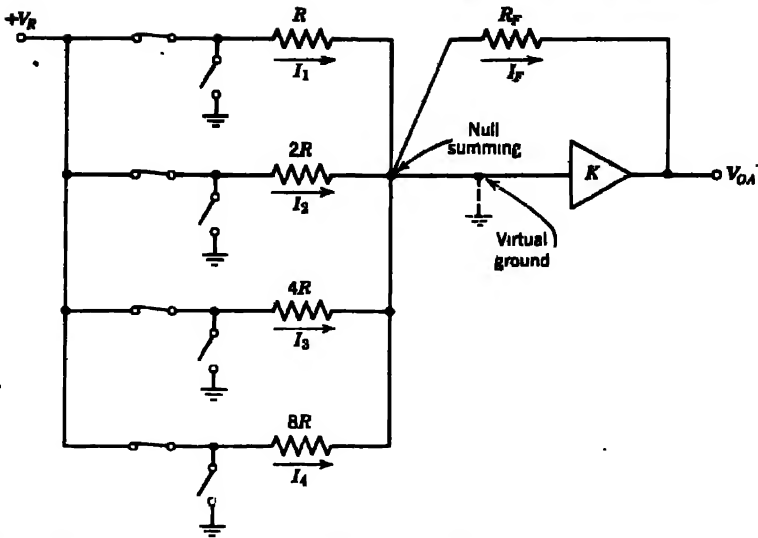


Figure 20. Null summing D/A decoder. (From Reference 2.)

An operational amplifier produces an output voltage 180° out of phase with the input; consequently

$$-V_{OA} = I_f R_f$$

$$-V_{OA} = \left(\frac{V_1 D_1}{R} + \frac{V_2 D_2}{2R} + \frac{V_3 D_3}{4R} + \cdots + \frac{V_n D_n}{2^{n-1}R} \right) R_f$$

The maximum full-scale output voltage when all switches are in the "1" state is as follows:

$$-V_{OFS} = \frac{V_R R_f}{R_n}$$

where V_{OFS} = full-scale output voltage.

The principal advantage of using operational amplifiers is that they provide very high input impedances, thereby ensuring that the input power to each branch is very low. The feedback resistor is also compatible with this approach. The output impedance of the amplifier is reasonably low so that it can be used conveniently with subsequent circuitry. It also provides good isolation between the basic circuitry and the load. The frequency response is good by present standards. The package size of operational amplifiers is quite small and they are also available as integrated circuits. The only limitation of this technique is possible amplifier oscillation, but this can be eliminated with standard design procedures.

1.5.1. D/A Design Parameters

Before investigating the merits of available hardware it is important to determine some of the parameters required by the proposed system.

Resolution is one of the most basic considerations. The maximum decoding error is $\pm \frac{1}{2}$ the value of the LSB; therefore the greater the number of bits, the better the resolution and accuracy. Conversely, the more bits, the more components required and the higher the cost. A practical compromise should equate resolution against cost for a particular system.

Accuracy is usually expressed as $0 \text{ V} \pm \text{mV}$ with all inputs on the "0" state at a particular temperature (usually 25°C). It is sometimes expressed as a percentage of full-scale accuracy, such as 0.015% of full scale for a binary format. It is different for various codes. When the decoder is the closed-loop type, the null voltage is specified. An additional part of the accuracy figure is the temperature coefficient; typical values are $\pm 0.001\%/^\circ\text{C}$ plus the drift of the reference voltage power supply.

Full-scale output is defined as the output voltage when all inputs are in the "1" state. Manufacturers of this equipment normally specify the load resistance, or output current, along with output voltage. If your system uses a nonstandard load resistor, the package can be needlessly expensive. Another parameter tied in with the output is the settling time for a full-scale voltage step. This is typically about $50 \mu\text{sec}$.

The load capacitance and estimated stray capacitance should be approximated if frequency response and high resolution are critical. Representative values are 0.1 to $0.2 \mu\text{F}$.

The proposed system must be designed so that it provides compatible supply voltages for existing hardware. Typical reference voltages are $\pm 15 \text{ V}$ DC with high stability and low drift. Current semiconductor technology produces full-scale outputs of about 10 V for inputs of 2.0 to 5.0 V at "1" state and 0 to 0.8 V for "0" states.

Nominal operating temperatures are $+10$ to $+50^\circ\text{C}$, but much wider tolerances are possible.

The available envelope for the converter is another parameter worth checking before you call the sales representative. The size of competitive units is constantly being decreased and is well worth scrutinizing.

Now that we have discussed the primary design parameters of D/A converters, let us examine some of the sources of the errors.

The reference voltage is one of the troublespots. Its output may vary with time and temperature. You can purchase references with overall regulation of 0.01 to 1% depending on your budget. In addition, it has a small but finite internal resistance, r_p . The current supplied to the network, I_N , and the current to the switch driver circuits, I_D , produce a voltage drop across r_p .

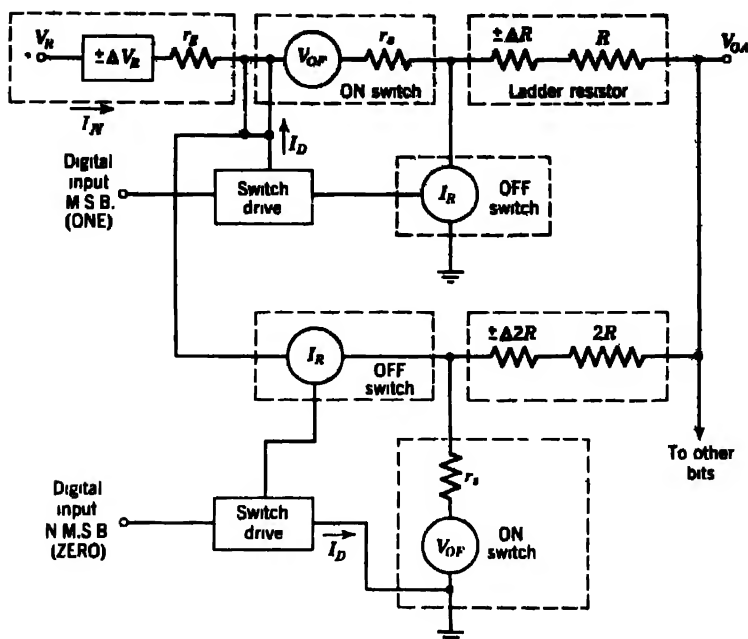


Figure 21. Equivalent circuit for typical D/A decoder. (From Reference 2.)

that introduces an error in the output of the network (Figure 21).

The switch used in the circuit is typically some kind of semiconductor. In the circuit shown it is driven by an auxiliary circuit that draws current from the reference voltage. The switching transistor has two inherent sources of errors: an offset voltage, V_{OF} , that may exist across the switch even in the absence of a load current I_L , and, a much less serious error, the leakage current through the transistor, I_R .

One of the prime sources of error in the converter is the resistors used in the networks. The largest error is the tolerance on the absolute value of resistance; this ranges from $\pm 5\%$ for carbon units to as low as $\pm 0.01\%$ for precision wirewound resistors. However, a good design uses only matched resistors, which can be held within $\pm 0.001\%$ of each other. The resistors also change value with age; this effect ranges from ± 0.5 to $\pm 0.002\%$ /year. The change in value due to thermal effects ranges from ± 1 to 500 ppm/ $^{\circ}\text{C}$. This is one category of components definitely worth the expenditure of money. Wirewound and metal film resistors are preferred for decoders.

1.6. SUMMARY

We examined A/D and D/A converters in Chapter 1 because the rest of the topics presented in this book should be viewed against a background of the inherent advantages of digital techniques. In short—"think digitally."

The main reason why digital components have evolved so slowly is that there is still a shortage of design talent that can handle the problems as proficiently as those presented by analog techniques. The number of electrical engineers schooled in digital techniques during the past 5 to 10 years is still only a small percentage of those designing, buying, and specifying systems. This condition will certainly be reversed during the next decade.

The second reason for the domination of analog devices is price; analog instruments are still cheaper than most digital approaches. However, with the development of mass-produced integrated circuit A/D converters the difference in price may become so small compared to the advantages that a fresh approach will be possible.

Another constraint on digital progress has been that most nontechnical people are used to seeing dials rather than digits and still expect to find them in the new models they purchase. This might be compared to attempts to convert our systems of weights and measures to the metric system; the change is coming but ever so slowly.

Human factors engineers also hold differing opinions about which class of device is visually better in case of an extreme emergency; they use the premise that most people are familiar with dial-type analog readouts. The standard altimeters of commercial aircraft reflect this dilemma: the altitude is indicated by a multipointer dial as well as by an odometer-type counter. The newer vertical-scale indicator presently being developed uses a tape of printed numerals that sweep by a stationary pointer. This is basically an analog presentation but a conventional digital altitude readout is also available.

Consider your own reaction to an automobile instrument panel consisting of a digital speedometer, oil pressure gage, water temperature gage, and fuel gage. Would the added accuracy be worth \$25 to \$50 more when you purchase the car? When you include the standard odometer and possibly a tachometer, would the array of digits present a problem when quick readings are necessary?

The material on sensors presented in Chapter 2 is filled with possible applications for digital readouts. Most manufacturers still provide data sheets that are only intended for analog systems. Pressure transducers are a notable exception because of the influence of military technology. Still, not all pressure transducers are adaptable to this approach, and there are numerous possibilities for improving these devices.

Encoders, discussed in Chapter 3, are the basic tool used for converting rotary motion into a train of digital pulses. Although many exotic types of encoders are described, the emphasis here is on extremely low cost units. The 19- and 20-bit encoders developed for military applications costing \$5000 and more can have little practical value in the automation industry and certainly none in the automotive field. The present state of the art is directed toward a good 7-bit (128 parts per revolution), \$50 unit. Even cheaper units must be developed before they can be used in the automotive and appliance industries. Perhaps basically simpler and less expensive designs must be conceived before encoders will rival commercial analog instruments.

The balance of the book deals with electromechanical components that all lend themselves to digital measurements and controls. It is hoped that the reader will view the devices discussed with these possibilities in mind.

Analog systems are far from passé. Even the most elegant digital system can profit from some analog outputs. Any system that frequently swings from a positive to a negative value is easier to monitor with a meter movement than a counter. When it is important to note the rate of change of a process variable, analog meters are easier to follow than digital units. Numerous examples of this are given in Chapter 4. Some systems simply do not warrant the expenditure of money required by digital outputs; for example, the gage pressure of a pneumatic controller, the vacuum reading on an exhaust pump, and other normally constant processes where readings are not critical.

In summary, this chapter should provide the necessary framework for properly viewing electromechanical components in new analog and digital systems.

REFERENCES

1. *Electromechanical Design*, January 1965, Benwill Publishing, Brookline, Mass.
2. D. E. Hoeschele, Jr., *Analog-To-Digital/Digital-To-Analog Conversion Techniques*, John Wiley & Sons, Inc., 1968.
3. *Analog-Digital Conversion Handbook*, E-5100, Digital Equipment Corporation, Maynard, Mass., 1965.
4. S. Wald, "A/D Conversion Techniques", *Electronic Design*, August 2, 1962, Hayden Book Company, New York.
5. J. A. Walston and J. R. Miller, Ed., *Transistor Circuit Design*, McGraw-Hill Book Company, New York, 1963.
6. N. S. Palazzini, *Hybrid Microcircuits for D-to-A Converters*, Sprague Electric Company, North Adams, Mass., 1966.

Chapter II · Sensors

The sensors used in analog and digital systems are essentially the same. As shown in Chapter I, the important difference lies in the readout technique. This chapter reviews the sensors normally used in industry. For each conventional sensor listed, there are probably two that have been developed for special applications. The complete list is boundless and will continue to grow as technology advances toward new horizons. Unfortunately, not every technique is commercially practical. This review is confined to available hardware.

Most sensors measure one of the following quantities (Figure 1):

1. Linear or angular displacement.
2. Velocity.
3. Acceleration.
4. Pressure.
5. Temperature.

2.1. LINEAR DISPLACEMENT TRANSDUCERS

System designers normally select displacement sensors that provide a linear or gradually increasing response to changes in displacement. The following terminology is used to define this characteristic (Figure 2):

Linearity: the percentage of deviation from a straight line when sensor output is plotted against displacement; for example, a 1% linearity is superior to a 10% linearity (sometimes called deviation).

Terminal linearity: the deviation when the slope and position of the line are established by the zero and maximum output points.

Independent linearity: the deviation when the line is selected over a narrower range than the full displacement of the transducer. This is permissible only when the sensor will function over a restricted range and exceptional linearity is required.

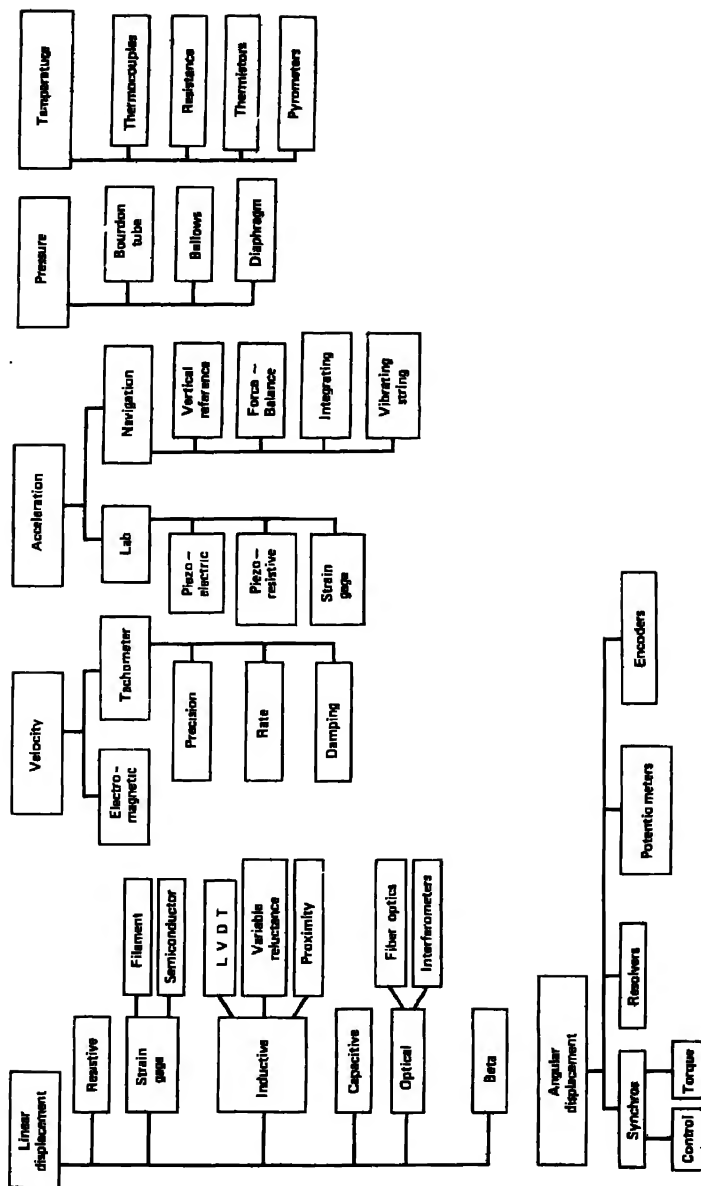


Figure 1. Classification of sensors.

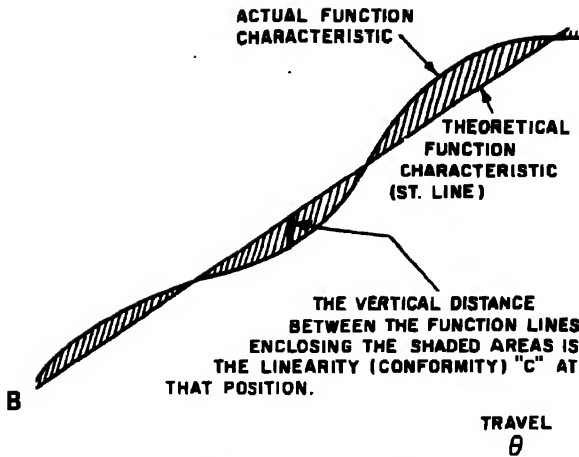


Figure 2. Linearity defined. A is given slope and B is given intercept at $\Theta = 0$. Unless otherwise specified, $A = 1$ and $B = 0$. (Courtesy of Benwill Publishing Company. From Reference 1.)

Slope: the change in output for a given displacement of the transducer over its operating range. This is typically expressed in volts per inch or ohms per centimeter.

Deviation: the plus or minus difference between the actual sensor output at a particular point and a straight line drawn between the zero and maximum output points. It is normally expressed in percent of maximum output.

Resolution: the smallest increment of change that can be detected by the sensor. In wirewound sensors it is the reciprocal of the number of turns of wire used in the instrument.

Linear displacement transducers are categorized as resistive, inductive, capacitive, or optical devices.

2.1.1. Resistive Transducers

The most commonly used resistive transducer is the potentiometer. Its most important usage is as a feedback element that converts position to an electrical error signal in servosystems (Figure 3).

A potentiometer is an electromechanical device containing a resistive element that is contacted by a movable slider. The position of the slider (or wiper) determines the output resistance of the device. The wiper is

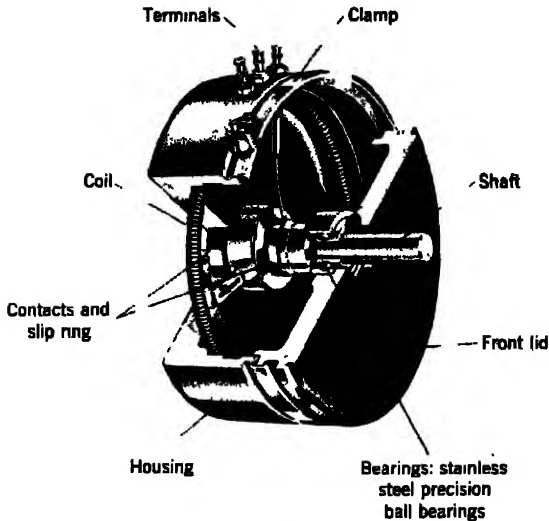


Figure 3. Precision-type potentiometer. (Courtesy of Beckman Instrument, Inc. From Reference 2.)

coupled to a mechanical element in the system so that a fixed relationship exists between the potentiometer output and the process being monitored. The simplest form of linear potentiometer consists of a length of high-resistance wire wrapped around an insulated bar, or mandrel. It contains three terminals, one at each end of the winding and a third connected to the wiper. Although this discussion is based on rotary potentiometers, the only difference between rotary and linear devices is a lead screw device that converts angular to linear motion.

There are five principal types of potentiometers on the market (Figure 4):

1. *Wirewound.* These use nickel-chromium, nickel-copper, or precious metal resistance elements. When properly heat-sunk they can carry relatively large currents at high temperatures. The temperature coefficient of resistance is low; typically 20 parts per million per degree Centigrade (20 ppm/°C) or less. Electrical noise characteristics are random in nature and deteriorate with age. Noise characteristics are determined by the surface finish of the winding and the wiper, the effects of dirt, dust, and humidity, and the resultant wear and oxidation of contact surfaces. Resolution is limited by the number of turns of wire on the mandrel.

2. *Cermet.* Cermet is a generic term used to describe a particular process that fuses precious metal particles onto a ceramic base. Its virtues are high

CHARACTERISTICS	WIREWOUND	NONWIREWOUND TRIMMERS			
		Cermet	Hot-Mold Carbon	Carbon Film	Thin-Metal Film
Setting Ability	Poor to Good	Excellent	Excellent	Excellent	Excellent
Resistance Range	Low to Medium	Low thru High	Medium to High	Medium to High	Low to Medium
Power Rating	Medium	High	Low	Low	Medium
Temp. Coefficient	Lowest	Low to Medium	High	High	Low
Environmental Stability					
High Temperature	Good to Exc.	Excellent	Poor	Fair	Excellent
Load Life	Good to Exc.	Excellent	Fair	Fair	Excellent
Humidity	Good to Exc.	Excellent	Poor	Poor	Excellent
Rotational Life	Good	Excellent	Good	Good	Fair
Rheostat Usage	Good	Fair	Poor	Poor	Fair
AC Usage	Fair	* Excellent	Excellent	Excellent	Excellent

Figure 4. Potentiometer characteristics. (Courtesy of Beckman Instrument, Inc. From Reference 2.)

power ratings at high temperatures plus low cost. Temperature coefficients are moderate: 100 to 200 ppm/°C.

3. *Hot-molded carbon.* A mixture of carbon and a thermosetting plastic binder that is fabricated by molding processes. Characteristics include noise parameters similar to wirewound units but considerably better AC features.

4. *Carbon film.* Thin deposits of carbon on a nonconductive base. One outstanding characteristic is low cost on noncritical applications. Temperature coefficient is up to 1000 ppm/°C.

5. *Thin metal film.* Very thin, vapor-deposited layer of metal on glass or ceramic base. Advantages include excellent resistance to environmental cycling and AC usage. Cost is moderate.

2.1.1.1. Selection of Potentiometers. Before selecting a potentiometer the following system parameters should be considered:

1. *Sealing.* Will the unit be in an environment that may cause corrosion to the contacting element? If so, specify a sealed potentiometer.

2. *Setting ability.* Many applications require that potentiometers be set very closely. The choice here is a nonwirewound unit, since the resolution on a wirewound instrument is limited by the size of the wire used. The resolution on Cermet and film-type units is "infinite." Multiturn potentiometers are very helpful when settings are critical; 3- and 10-turn units are commercially available.

3. *Power rating.* When power ratings are high, the choice should be limited to wirewound or Cermet units. If the load is AC, Cermet should be chosen. Film "pots" should be used only for trimmer applications. For maximum power ratings on any unit, the thermal path must be carefully designed. A good heat sink will provide the best assurance of long, reliable operation at rated and overrated conditions.

4. *Electrical noise.* The noise spectrum in wirewound units is random and nonrepeatable and usually changes during the life of the potentiometer because of contamination, wear, and oxidation. Film-type "pot" noise is repeatable and usually does not vary with time.

5. *Temperature coefficient.* An increase in resistance with temperature rise is usually considered a parasitic effect, and equipment should be selected to minimize it. The wirewound unit excels in this characteristic. The following equation defines the problem:

$$R_2 = R_1(1 + \alpha \Delta t) \quad (1)$$

R_2 = resistance at an elevated temperature

R_1 = initial resistance (cold)

α = change in resistance per °C; typically $20 \times 10^{-6}/^\circ\text{C}$ for wirewound units and up to 1000×10^{-6} for carbon "pots"

Δt = change in temperature in °C

6. *AC usage.* The interwinding capacitance between turns on a wire-wound potentiometer and between the winding and mandrel, limits the AC usage to low frequencies because of the increase in impedance. Non-wirewound units are almost free of this problem.

2.1.1.2. Failure Mode. To produce low contact resistance, the contact pressure between the wiper and resistance element should be as high as possible; to produce low friction levels, the contact pressure should be as low as possible. The normal compromise is a pressure of about 5 to 10 grams. When lighter loads are used, the cleaning action of the wiper is ineffective, and normal airborne contaminants may produce high contact resistance. The noise level also increases. When the unit is subjected to shock and vibration, a low preload or contact pressure permits contact bounce. The wiper assembly is usually designed so that its natural frequency is greater than any anticipated vibration forcing frequency. Although low contact pressure should produce long life for mechanical reasons, the associated arcing normally results in pitting that reduces the life and increases the associated noise. The failure mode of potentiometers is increasingly excessive noise; therefore the signal-to-noise ratio is no longer useful for control purposes.

More detailed information on potentiometers is provided in Chapter 6.

2.1.2. Strain Gages

Both strain gages and potentiometers are resistive transducers; the basic difference is that potentiometer resistance is varied manually while strain gages exhibit a change in resistance caused by mechanical deformation, or strain (Figure 5).

Strain gages are slender metallic filaments attached to a thin base material that is in turn fastened to the surface being monitored. When the surface is deformed because of applied load, the strain gage is also deformed; this causes a change in resistance. The minute change in resistance is detected by a Wheatstone bridge and related mathematically to the applied load.

Occasionally strain gages are used that consist simply of a wire wound on the test specimen. They are referred to as unbounded gages and are considerably more difficult to use than the bonded type.

2.1.2.1. Types of Measurements. Measurement of static strains is one of the most common applications of strain gage sensors. Typical examples are aerodynamic studies, structural fatigue investigations, and marine programs (Figure 6). Frequency response may vary from zero to 50 kHz.

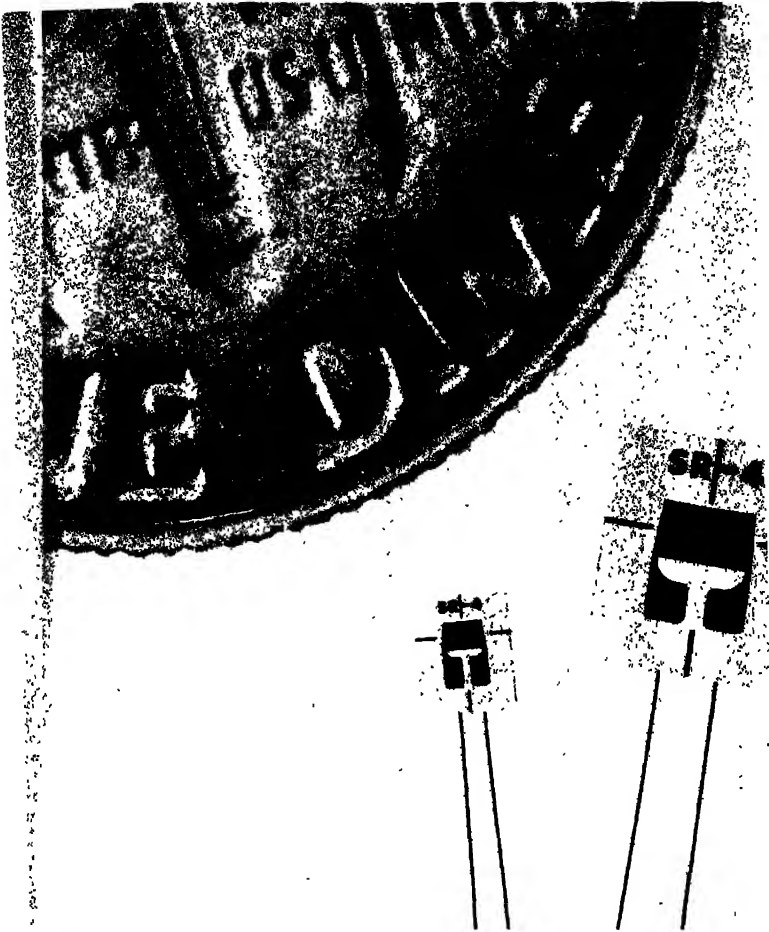


Figure 5. Strain gage (10 times actual size). (Courtesy of BLH Electronics, Inc. From Reference 3.)

Gages should be selected to provide the greatest response in the sensitive axis; this is normally along the longitudinal axis of the instrument. Conversely, sensitivity should be minimal along the transverse axis. Minimum sensitivity to long-term thermal effects and dimensional stability are also important.

Dynamic strain measurements generally require gages of maximum strain sensitivity with less emphasis on stability. Typical examples are

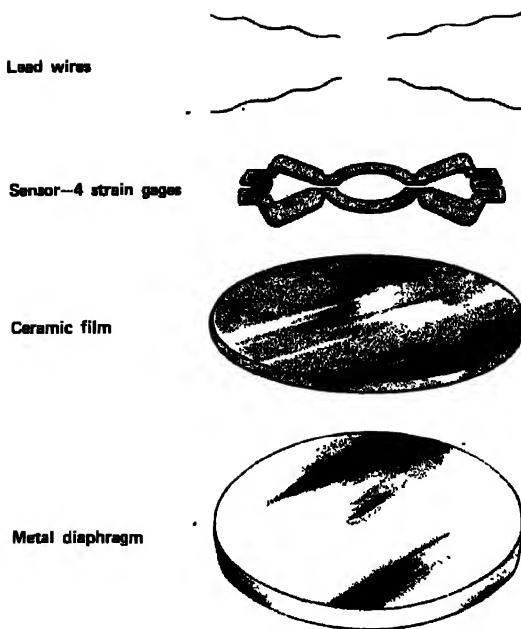


Figure 6. Strain gage transducer. The strain gage bridge consists of four active legs which have the same length/width ratio and are balanced internally to ensure electrical uniformity. Pressure is measured by the resistance change in the legs of the bridge. Applied pressure causes tensile stress in one pair of legs and compressive stress in the opposite pair of legs. The gages are applied by vacuum deposition techniques. (Courtesy of Statham Instruments, Inc.)

monitoring of shock phenomena, blast effect on structures, and the effect of rough roads on autos.

When the direction of the applied strain cannot be determined in advance, a three-element strain gage may be used. The three elements act independently and resolve the forcing function into three vectors. If individual units are mounted so that they are mutually perpendicular, the absolute magnitude and direction of the strain can be computed from the data recorded.

2.1.2.2. Range. Most wirewound strain gages are limited to strains that range from $\frac{1}{2}$ to 10% of their length. The most common range is $1\frac{1}{2}$ to 2%. This is determined largely by the wire used in the strain gage. The following example illustrates how the data can be applied.

Assume that we select a strain gage with a nominal resistance of 120 Ω and a gage factor* of 2 and that we mount it on a 10-in.-long steel rod.

* See Section 2.1.2.4 for the definition of gage factor.

The gage is $\frac{1}{2}$ in. long. The change in resistance, as measured on a Wheatstone bridge, is 0.5Ω . What is the indicated stress and the maximum stress that can be measured?

$$\begin{aligned} S &= E\varepsilon \\ \varepsilon &= \frac{\Delta L}{L} = \frac{1}{K} \frac{\Delta R}{R} \\ \therefore S &= E \left(\frac{1}{K} \frac{\Delta R}{R} \right) \end{aligned} \quad (2)$$

$$S = 30 \times 10^6 \left(\frac{1}{2} \times \frac{0.5}{120} \right) = 62,500 \text{ psi}$$

where S = applied stress

L = gage length (0.5 in.)

ε = measured strain

K = gage factor (2)

E = Young's modulus (30×10^6 psi)

R = gage resistance (120 Ω)

ΔR = change in gage resistance (0.5 Ω)

The maximum stress that can be measured, S_{\max} , is found as follows:

$$\begin{aligned} S_{\max} &= E\varepsilon \\ \varepsilon &= \frac{\Delta L}{L} = 0.01 \end{aligned}$$

where ΔL = change in length of the gage. This assumes that the maximum deflection of the gage is limited to 1% of the gage length.

$$S_{\max} = 30 \times 10^6 \times 0.01 = 300,000 \text{ psi}$$

2.1.2.3. Current Capacity. The amount of current that may safely be used in a strain gage is a function of the wire resistivity, the grid design, backing material, bonding current, and, most important, the heat sink to which they are attached. Gages bonded to aluminum or steel parts can tolerate much more abuse than those mounted on insulating materials. Paper-type gages bonded to metals are limited to 0.025 A while Bakelite types are rated as high as 0.050 A. Currents of 0.005 to 0.006 A are recommended on poor conductors such as plastics.

2.1.2.4. Selection. When selecting a strain gage, five basic variables must be considered:

1. The filament wire material.
2. The filament construction.

	ORGANIC						CERAMIC		
	Solvent Release		Chemical Setting				Chemical Setting	Fusing	
	Room Temperature Curing		Room Temperature Curing		Heat Curing		Heat Curing	Flame Spray	
BASE	Nitro-Cellulose		Acrylic	Epoxy	Epoxy	Phenolic	Phosphate	Refractory Oxide	
CEMENTS AND PROPERTIES	SR-4	Duco	Eastman 910	EPY-150	EPY-400 EPY-500 EPY-600	Bakelite	PBX	Rohide-BLH	
STRAIN GAGE COMPATIBILITY	Use with quick drying thin paper backed gages	All paper backed gages	All except paper wrap-around construction	All	All that will stand cure temperature Exception: All paper backed	Phenolic backed only	Strippable, transferable free-filament	Transferable, free-filament wire only	
SPECIMEN MATERIAL COMPATIBILITY	All except plastics soluble in MEK and Acetone and unbondable plastics	All except plastics soluble in MEK and Acetone and unbondable plastics	All except some plastics	All except some plastics	All except some plastics and reactive metals	All except some plastics and reactive metals	All except some plastics and reactive metals	All except some plastics	
CURE TEMPERATURE °F	Room to 150	Room to 150	Room	Room to 150	250-600	250-350	600	None	
CURE TIME	2-10 hours	12-48 hours	1-5 minutes	1-70 hours	1-30 hours	5-6 hours	1 to 6 hours	None	
CURE PRESSURE PSI	Contact 1-5 for 1 minute	1-5	Contact 1-15	5-15	15-30	50-100	None	None	

MAXIMUM OPERATING TEMPERATURE °F	180	150	150	150	400 normal, 500 after proper cure to 600 short time	300 continu- ous to 500 for short time	> 1000	> 1500
STRAIN LIMIT	> 10% at room	> 10%	> 10% at room	> 10% at room	> 5% at room > 1% at -320	> 2% at room > 1/2% at -320	1/4 %	1%
ELECTRICAL PROPERTIES	Excellent over operating temperature range	Excellent over operating temperature range	Excellent	Excellent	Excellent	Excellent to 300° F Poor in 400-500°F range	Deteriorates with increase in temperature above 1200	Excellent but deteriorates at high temperature above 1500
HUMIDITY RESISTANCE	Fair, absorbs up to 2% water	Fair, absorbs up to 2% water	Fair, absorbs up to 0.3% water	Good, absorbs up to 0.1% water	Good, absorbs up to 0.1% water	Fair, absorbs up to 0.2% water	Poor, is hygroscopic	Poor, is porous and hygroscopic
MOISTURE PROOFING RECOMMENDATIONS	Low RH, short test-time, Micro-Wax, Cerese, Dipel 171, all others	Low RH, short test-time, Micro-Wax, Cerese, Dipel 171, all others	Low RH, short test-time, Micro-Wax, Cerese, Dipel 171, all others	Normal RH, except paper gages-none Micro-Wax, all others	Normal RH, none, all others use Silicone resins, var- nishes or RTV Silastic	Normal RH, none, all others use Silicone resins, var- nishes or RTV Silastic	Conventional protection to 600°F, replace upon return room temp. or SD Sealant	Conventional protection to 600°F, replace upon return room temp. or SD Sealant
GENERAL APPLICATION REMARKS	Easiest, most reliable for large-scale testing	Good, gen- eral purpose cement with adequate drying	Fastest for short tests, poor thermal mechanical resistance	Easy and durable for long-term tests, normal ambients	Best long- term stability, varying ambi- ent conditions and transducers	Some prefer- ence for transducers but generally replaced by epoxies	Easier appli- cation with free-flament wire and foil gages	Fastest application, no appre- ciable heat on test specimen, rough sur- face recom- ended

Figure 7. Strain gage cement chart. (Courtesy of BLH Electronics, Inc. From Reference 3.)

3. The base (carrier) material.
4. Method of bonding the filament to the carrier.
5. Lead wire connection.

FILAMENT MATERIAL. Material is the major variable. Since the quantity actually measured is the change in resistance, it is desirable that the ratio of resistance change to strain be as high as possible. This ratio is called the gage factor:

$$\text{gage factor} = \frac{\Delta R/R}{\Delta L/L} = K$$

where $\Delta R/R$ = the ratio of resistance change to resistance

$\Delta L/L$ = the ratio of length change to length

If this effect were due entirely to dimensional change, we would expect, from Poisson's ratio, that the gage factor of any wire would be approximately 1.6. However, materials vary widely in this respect.

Constantan is a copper-nickel alloy used primarily for static strain measurements because of its low temperature coefficient and dimensional stability. Gage factor is about 2. Its temperature range is -100 to 300°F .

Nichrome V is a nickel-chrome alloy used at temperatures up to 1200°F for static measurement and 1800°F for dynamic tests.

Isoelastic alloys are used at temperatures up to 250°F where a higher gage factor is useful ($K = 3.5$).

Modified Karma provides a wider temperature compensation range than Constantan and is usable up to 600°F .

Platinum alloys are used for static tests up to 1200°F and for dynamic measurements up to 1500°F .

Semiconductor strain gages, to be discussed later, have gage factors of 4.5 to 200.

FILAMENT CONSTRUCTION. As mentioned previously, the two primary types of gage construction are unbonded and bonded units. Unbonded units are used where conventional bonded gages are not suitable because of the temperature limitations of the insulation material or geometric problems associated with the unit being tested. Bonded gages are in turn divided into wire and foil types. The wire type, which is the older type, is mounted on some form of carrier such as impregnated paper, bakelite, or epoxy, which is in turn bonded or cemented to the test unit. The filaments are wound in a grid pattern to minimize the length of wire that is perpendicular to the sensitive axis. Foil or thin-film gages are made by techniques similar to electronic microcircuitry. The resistance elements are vacuum-coated with a ceramic film and deposited on some form of backing. In some cases the backing is a strippable vinyl material that can be peeled off so as to cement the filament directly on the test unit with a ceramic adhesive. Foil gages have unusually low cross-axis sensitivity.

CARRIER MATERIAL. Several types of base or carrier material are used to support the filament in filament-type gages. For room temperature applications, impregnated paper is used. For slightly higher temperatures, an epoxy is used. This provides a useful range of -320 to 250°F . Bakelite-impregnated cellulose or glass-fiber filled materials are employed for continuous service up to 350°F and for limited operation at 500°F .

ADHESIVES. The cements used for fastening gages to the test unit are summed up in Figure 7.

LEADS. It is necessary to select materials that have low, stable resistivity and minimum temperature coefficients. They must be insulated with materials that are at least as good as the gage backing (Figure 8). The simple soldered junction between the filament and lead wires is satisfactory only for static or slowly varying loads. When gages are to be used in fatigue studies, or when large strains are involved, dual lead wires are recommended.

2.1.2.5. Temperature Compensation. In addition to the five basic variables, temperature compensation must be carefully considered. The "active dummy" method uses an active gage bonded to the test specimen and a second identical gage (dummy) bonded to a separate part of the same material; the dummy is not subject to strain, but is in the same thermal environment as the active gage. In the unstressed state both gages have the same response to temperature change. Both are connected into adjacent arms of a Wheatstone bridge where the temperature effect is canceled but the strain reading on the active gage is unaffected.

Another method is the specification of a selected melt of Constantan or Karma alloy where the temperature coefficient can be held to $\pm 1 \mu\Omega/\text{in.}/^{\circ}\text{F}$ over a temperature range of $+75$ to 600°F for Karma and $+50$ to 250°F for constantan.

Conductors	Operating Temperature $^{\circ}\text{F}$	
	Stable	Maximum
Nickel-clad copper	700	1000
Stainless steel-clad copper	800	1300
Nickel-clad silver	1000	1500
Nichrome	700	1700
Insulation	Temperature Range	
	Below -100°F	
Nylon	-100 to 150°F	
Vinyl	-100 to 200°F	
Polyethylene	-100 to 500°F	
Teflon	above 500°F	
Glass impregnated Silicone	above 500°F	
Glass sleeving	above 500°F	

Figure 8. Operating temperatures of conductors and insulation used with strain gages. (Courtesy of BLH Electronics, Inc. From Reference 3.)

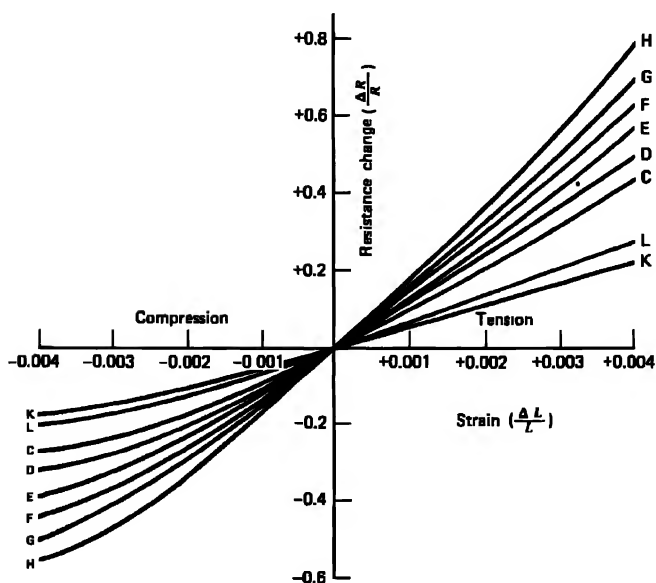
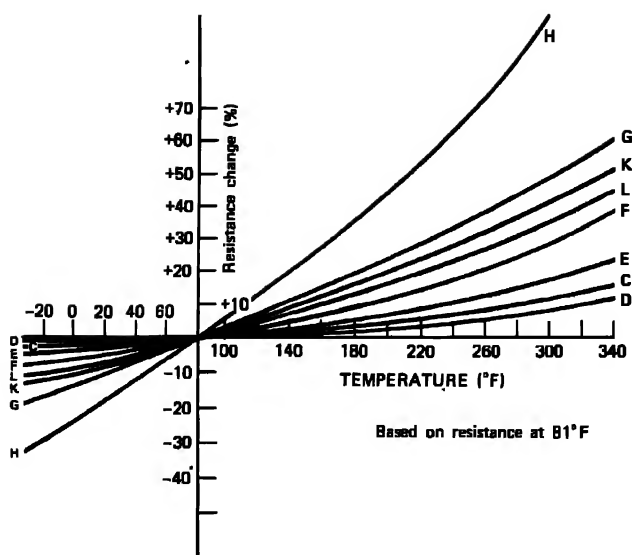


Figure 9. Characteristics of semiconductor strain gages. The letters H, G, F, E, D, C, L, and K represent particular gages. (Courtesy of Kulite Corporation. From Reference 6.)

The dual-element method of temperature compensation consists of using two constantan wire elements in series; one element has a positive temperature coefficient and the other a negative coefficient. These gages are designed to be applied to specific materials such as steel, titanium, or aluminum and the temperature coefficient is adjusted accordingly. Over a temperature range of $+50$ to 250°F the temperature coefficient tolerance may be limited to $\pm 0.25 \mu\text{in./in./}^{\circ}\text{F}$.

2.1.2.6. Semiconductor Strain Gages. Semiconductor strain gages are used where very high gage factors and small envelopes are required. Like metallic strain gages, they change resistance as a function of applied strain. A typical strain gage consists of a strain-sensitive crystal filament and leads that are sandwiched in a protective matrix. Gage factors of 4.5 to 200 are typical values. When selecting semiconductor strain gages, three critical engineering parameters must be considered (Figure 9):

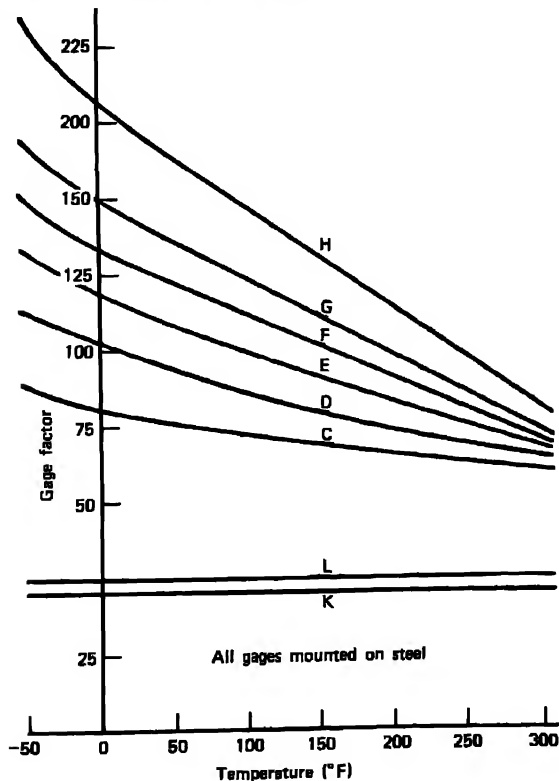


Figure 9.—Continued.

1. *Gage factor.* The higher the gage factor, the poorer the linearity.
2. *Temperature coefficient of resistivity.* The TCR may vary from about 2%/100°F to 50%/100°F.
3. *Temperature coefficient of gage factor.* The TCGF may vary from 0%/100°F to 20%/100°F.

The lower the resistivity, the better the linearity. Hysteresis characteristics of semiconductor gages are excellent; some units maintain less than 0.05%. Fatigue life is in excess of 10×10^6 Hz and frequency response is up to 10^{12} cycles. Temperature compensation techniques are similar to those used with metallic strain gages.

2.1.2.7. Circuitry. The most commonly used circuit is the Wheatstone bridge (Figure 10). The strain gage can be employed as one of the resistors, such as R_A , and the output voltage monitored as a function of the change of its resistance. For a Wheatstone bridge the following equation is applicable:

$$E = V \left(\frac{R_A}{R_A + R_C} - \frac{R_B}{R_B + R_D} \right)$$

A more convenient form of this equation is the following:

$$\Delta E = \frac{R_A R_D K \epsilon V}{[R_A + R_D]^2} \quad (3)$$

where ΔE = change in output voltage

K = strain gage factor

ϵ = strain

V = supply voltage

Other symbols are shown in Figure 10.

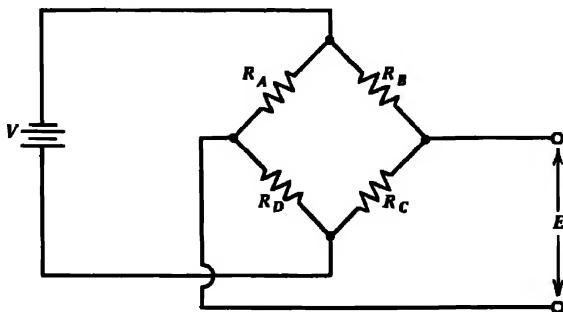


Figure 10. Wheatstone bridge circuit. R_A and R_D are strain gages, R_B and R_C are balancing resistors, V is supply voltage, and E is output voltage.

This equation is normally used with the bridge in the "unbalanced" mode. By "unbalanced" we mean that the output of the device is not at null. Most strain gage measurements are made in this way. If we waited to null the device, as when calibrating resistors, it would be impossible to record the dynamic nature of strain gage measurements.

Commercial strain gage bridges normally use an AC bridge power supply and subsequently demodulate and amplify the output so that greater sensitivity can be provided. The normal excitation frequency is 1000 Hz, providing frequency response up to about 100 Hz. Since the nominal frequency response is about 10% of the carrier frequency, problems requiring better response will necessitate higher carrier frequencies. Instruments utilizing 20 kHz excitation frequencies are commercially available (Figure 11); they usually have their own recording oscillograph with provision for viewing the waveform on an external oscilloscope.

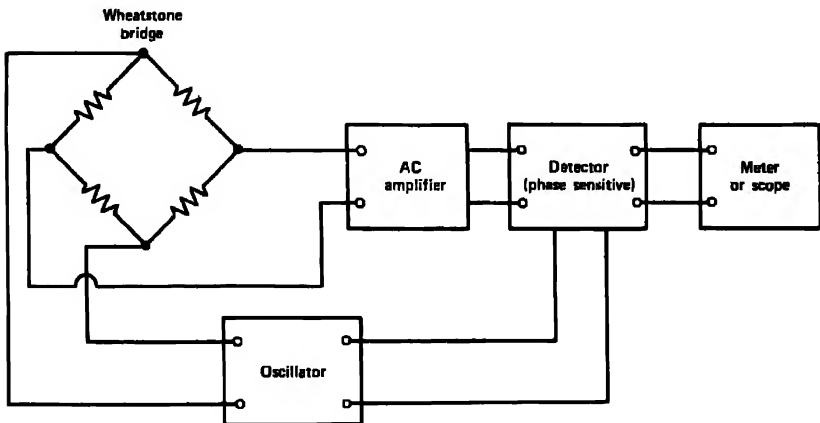
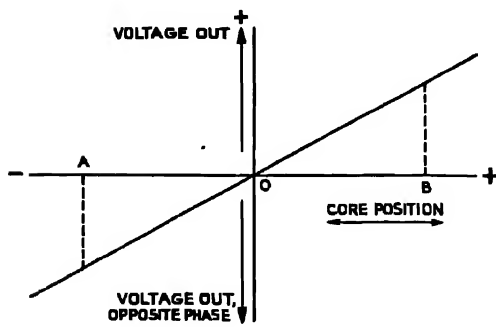
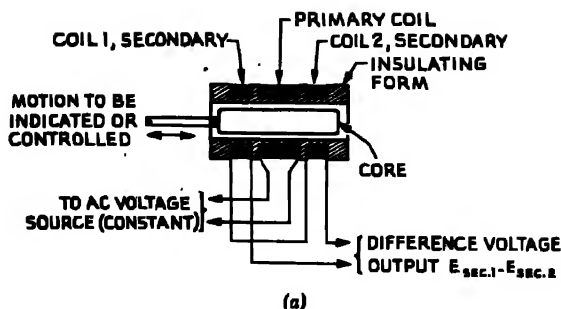


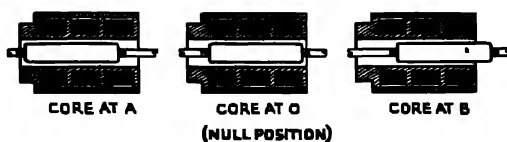
Figure 11. Commercial strain gage instrumentation. System oscillator operates at frequencies of 1 to 20 kHz.

2.2. INDUCTIVE TRANSDUCERS

The most widely used inductive transducer is the linear variable differential transformer (LVDT). This electromechanical transducer produces an electrical output proportional to the displacement of a movable core (Figure 12). It consists of three windings equally spaced on a cylindrical coil; the control winding is the primary and the two adjacent coils are the secondaries. A rod made of ferromagnetic material positioned axially inside the coil



CORE DISPLACEMENT



(b)

Figure 12. LVDT operating characteristics. (a) Operation of the linear variable differential transformer (LVDT). (b) LVDT output voltage and phase as function of core position. Courtesy of Schaevitz Engineering Company. From Reference 8.)

provides a magnetic path between the three coils. When the primary winding is energized with alternating current, a voltage is induced in each of the secondary windings. The secondary windings are connected in series opposition, so that the two voltages generated are 180° out of phase and the net output of both windings is the difference between the two voltages. When the core is equally spaced between the two windings, the device is in

its "null" position; the net output voltage is zero. As the core is moved from null to a point closer to one of the windings, the effective air gap between the primary and that secondary winding decreases, the reluctance of the magnet circuit linking them decreases and the flux as well as the net output voltage developed increase. Conversely, the voltage in the other secondary winding decreases. Since the two secondaries are connected in series opposition, the net output of the LVDT reflects the phase of the dominant secondary.

The advantages of the LVDT are as follows:

1. Infinite resolution—the change of voltage is "stepless." There are no mechanical elements to change voltage in discrete steps, as in potentiometers.
2. High sensitivity—as high as 1 V/0.001 in.
3. Good linearity—0.05% linearity is commercially available.
4. Ruggedness—usually can tolerate a high degree of shock and vibration without degradation of performance.
5. Low hysteresis—repeatability is excellent under all conditions.

2.2.1. Linearity

The output voltage of an LVDT is a linear function of core displacement within a limited range of motion. Beyond this range the curve starts to deviate from a straight line. In current literature the term linear range, or full range, refers to the total linear displacement in both directions from the null point of the LVDT. Full-scale displacement refers to the travel from null to either limit of the linear range. Linearity is normally computed by moving the core in a number of accurately measured increments through the linear range and observing the electrical output for each point. The definition of linearity as recommended by the Instrument Society of America involves computation of the standard deviation from a set of points. The applicable equations are as follows:

$$\sigma = \sqrt{[\sum(Y_1 - Y)^2]/(n - 1)} \quad (4)$$

where σ = standard deviation

Y_1 = observed value of output for a specified position X_1

Y = theoretical value of output for position X_1

n = number of observed pairs of X_1 and Y_1

A best estimate of linearity is

$$\% \text{ linearity} = \frac{0.67\sigma}{\text{linear range}} \times 100$$

Very often the linearity is defined as a percentage of the output at nominal range. For example, if the output of a LVDT is 1.50 V at maximum travel, and the maximum deviation of the output curve from a straight line through the origin is ± 0.003 V, the linearity is

$$\frac{\pm 0.003}{1.50} \quad (5)$$

Data on linearity and linear range are always given in conjunction with a specified load; 0.5 M Ω is commonly used.

2.2.2. Sensitivity

The rated sensitivity of an LVDT is usually stated in terms of millivolts of output per 0.001-in. core displacement per volt of bridge excitation; an alternate measurement is in volts per inch per volt. The frequency of the excitation must also be stated, since sensitivity is a function of frequency (Figure 13).

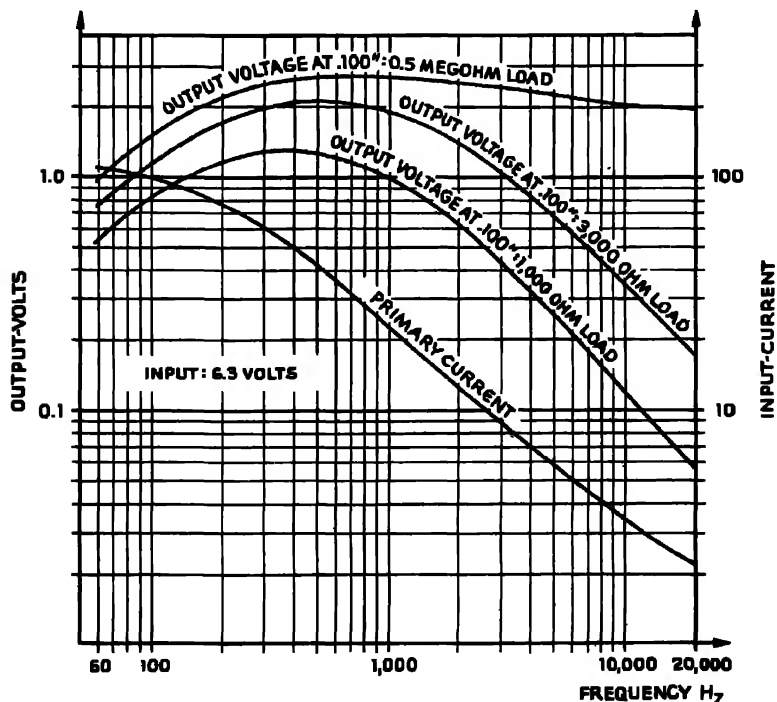


Figure 13. Sensitivity of an LVDT versus frequency. (Courtesy of Schaevitz Engineering Company. From Reference 8.)

2.2.3. Load Impedance

Although the LVDT is inherently low in output power, it has sufficient power to drive some motors or the control winding of a magnetic amplifier. As in any other electrical network, maximum power delivery is achieved by selecting the load impedance equal to the output impedance of the LVDT. When the load is fixed, it is often possible to purchase an LVDT with a nonstandard output impedance; this is accomplished by using different wire sizes in the instrument.

2.2.4. Resolution

The output voltage characteristic of the LVDT is "stepless." The effective resolution depends more on the test equipment than the transducer. Normal instrument practice results in a resolution of 0.1% or better.

2.2.5. Phase Characteristics

The phase angle of the output voltage with respect to the input voltage has two values differing by 180° ; they correspond to the position of the core on either side of the null position. When not explicitly stated, the phase angle referred to is the one closer to zero. The phase angle generally is between -20 and $+75^\circ$, depending on frequency, load, and other factors (Figures 14 and 15). To calculate the phase angle of the primary current with respect to the input voltage, the following approximate relationship is used:

$$\tan \phi = -2\pi f L_p / R_p \quad (6)$$

$\phi = \text{phase angle}$

The voltage in the secondary leads the primary current by about 90° ; therefore the overall phase shift can be calculated. Normally as the core

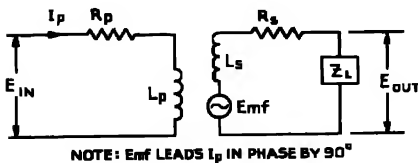


Figure 14. Equivalent circuit of LVDT. R_S = Secondary resistance, L_S = secondary inductance, Z_L = load impedance, I_P = primary current, R_P = primary resistance, and L_P = primary inductance. (Courtesy of Scharvitz Engineering Company. From Reference 8.)

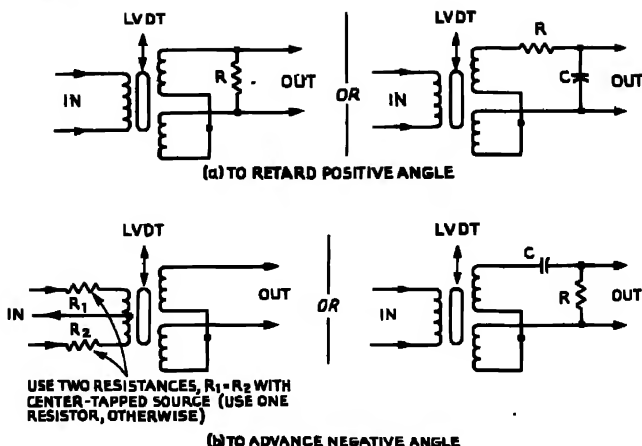


Figure 15. Practical circuits for reduction of LVDT phase angle. (Courtesy of Schaevitz Engineering Company. From Reference 8.)

passes through the null position the output phase changes abruptly by 180° . Under conditions of high null voltage, the reversal may be less abrupt.

2.2.6. Null Voltage

The null voltage is composed of three components: quadrature voltage, harmonics, and noise. A minimum null voltage results when the amplitudes of the two secondaries are equal. The difference in phase angle between the output voltages of the secondaries should be approximately zero. However, nonsymmetrical windings, nonuniform wiring, and unsymmetrical magnetic circuits cause slight phase angle differences. A practical solution is to place a potentiometer or variable capacitor across one of the windings and adjust it until the phase shift is zero.

The third harmonic is the major component of the harmonics present in the output. The magnetic materials in the transducer are the basic source of harmonics. The null position of the harmonics is located at a different point from that of the fundamental. One way to minimize the effect is to feed the input voltage, reversed in phase, into the secondary circuit. The opposite-phase voltage cancels out the components existing at the third harmonic null position. Another solution is to feed the output through a low-pass filter.

The noise content of the null voltage is due to the irregularities in the magnetization of the magnetic circuit. A small capacitive load in the output generally reduces the noise.

2.2.7. Temperature Characteristics

The temperature sensitivity of an LVDT is primarily a function of the resistance and inductance. At low frequencies the change in resistance is the dominant effect. A change in resistance alters the current and the flux generated; the net effect is a corresponding change in sensitivity. At high frequencies inductive reactance, which is unaffected by temperature, becomes the predominant part of the impedance, and resistance variations have less effect. Typically, the change in output sensitivity at low frequencies is about 10 to 12% for a temperature increase of 100°F. For every LVDT an optimum excitation frequency exists where losses and gains balance and produce no change in sensitivity due to thermal effects. An increase in temperature also causes the phase angle to shift in the positive direction. This can be minimized by placing a capacitive load across one of the secondary windings. Null shifts due to differences in thermal coefficients of expansion are difficult to eliminate. The stator, composed largely of copper, and the core, made of a highly permeable steel, expand at different rates; the net result is a null shift at high temperatures. Fortunately, this has only a small effect on sensitivity.

When temperature effects are expected to affect system performance seriously, manganin wire may be substituted for copper; the temperature variation is thereby reduced to almost zero. Unfortunately, manganin has a high resistivity and reduces sensitivity to about one fifth of that of comparable copper coils. At higher frequencies, where inductive reactance becomes the dominant effect, manganin may be used more efficiently.

2.2.8. Proximity Inductive Transducers

Although the vast majority of linear inductive transducers in use today are LVDTs, two additional instruments are occasionally useful: the mutual inductance pickup and the variable reluctance pickup.

The mutual inductance pickup consists of primary and secondary coils wound on a form (Figure 16a). The primary is excited by an AC source. When a ferromagnetic object is brought near the field generated by the primary, the flux linking the two coils is increased and the output from the secondary also increases. The voltage generated in the secondary is expressed as follows:

$$V = M\omega I_p \quad (7)$$

where M = mutual inductance (henries)

ω = exciting frequency (radians per second)

I_p = the current in the primary (amperes)

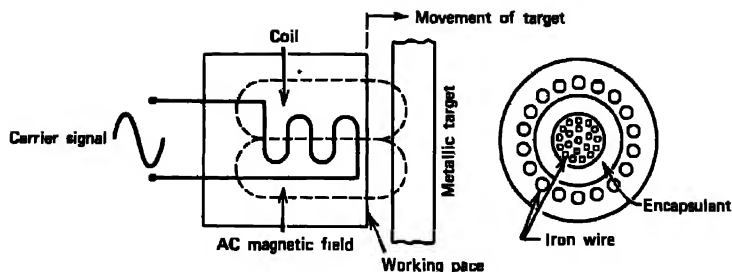


Figure 16a. Principle of operation of proximity transducers—the primary coil, supplying the magnetic field, is in the center of the instrument. The secondary coil, or output winding, is wound around the periphery of the device. When a ferromagnetic object is brought near the field generated by the primary, the flux linking the two coils increases and the output from the secondary increases proportionately. (Courtesy of MTI Corporation. From Reference 9.)

Motion of the ferromagnetic object or target modulates the output of the secondary and the result is an amplitude-modulated signal that is subsequently amplified, demodulated, and read on a scope or strip chart recorder. The relationship between the motion of the target and the output of the transducer is linear for a relatively small excursion and varies with the shape of the core. The big advantage of the technique is that it results in extremely small sensors that measure position without physical contact (Figure 16b). The instrument is not affected by most contaminants and will function at temperatures over 1000°F. The chief limitations are its nonlinear output and the requirement that the target be a ferromagnetic material.

Variable-reluctance transducers may be considered the forerunners of



Figure 16b. Typical size of Intronic proximity transducer (MTI trademark). (Courtesy of MTI Corporation. From Reference 9.)

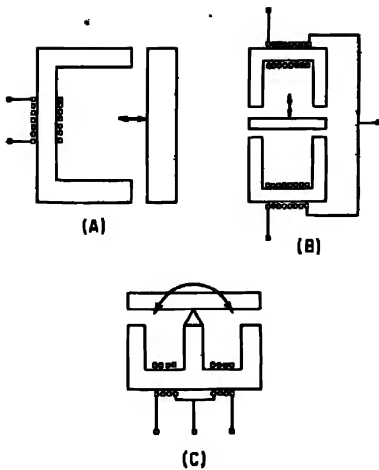


Figure 17. Variable-reluctance principle air-gap type. The inductance of the coil is changed by varying the air gap between the movable and stationary parts of the magnets. Designs including (A) a single magnet, (B) a double magnet, and (C) a E-core magnet are used. The coils are usually arranged as part of a Wheatstone bridge circuit. (Courtesy of McGraw-Hill Book Company. From Reference 7.)

the modern LVDT (Figure 17). The air gap is very small compared to the total length of magnetic core material. For small air gaps the change in inductance of the coil is linearly proportional to the change in air gap. Small motion of the magnetic target results in a linear output of the device. The two-coil type is used to form two arms of a Wheatstone bridge. The output is an amplitude-modulated signal that is amplified, demodulated, and viewed on a scope. Typical air gaps are under 0.035 in.

2.3. CAPACITIVE TRANSDUCERS

Capacitive transducers are used when exceptionally high sensitivity to displacement is required; measurements down to the microinch range are typical. The principle of operation is based on the familiar capacitance equation for a parallel-plate capacitor in air:

$$C = \frac{KA}{S} \quad (8)$$

where C = capacitance (microfarads)

K = dielectric constant for the medium between the plates

A = effective area of the plates (square centimeters)

S = separation between plates (centimeters)

When extremely fine displacements are to be measured, the distance between plates of the transducer is the variable. This is physically accomplished by suspending one plate over the object being monitored (Figure 18). The object forms the second plate.

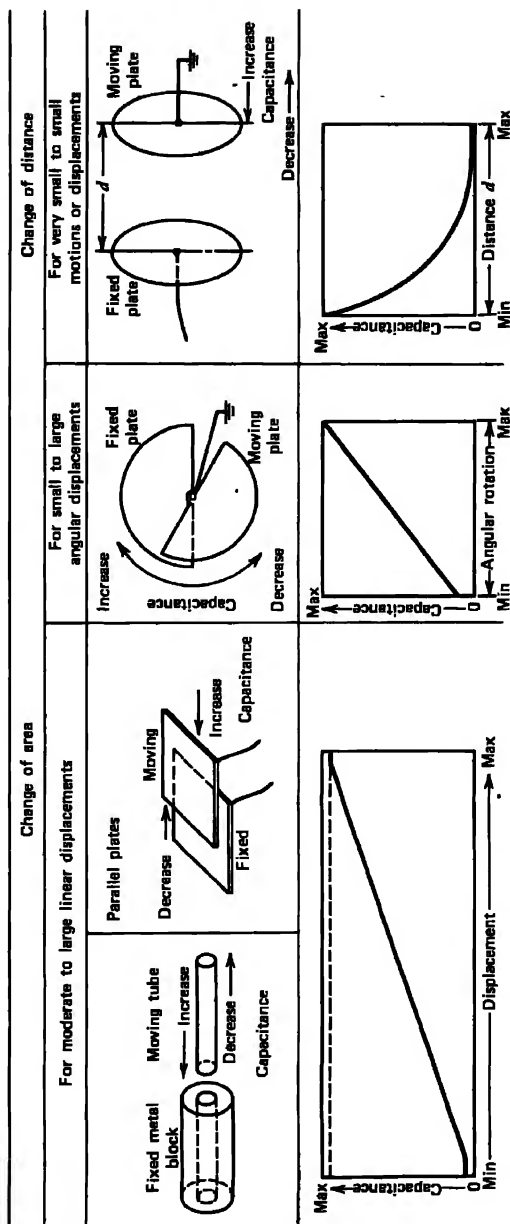


Figure 18. Capacitance transducers for linear or angular displacements. Basic transducer types fall into two categories: those that operate by changes in area and those that operate by changes in distance between plates. (Courtesy of McGraw-Hill Book Company. From Reference 10.)

If the motion to be studied is in the range of $\frac{1}{16}$ to several inches, it is more practical to vary the area between plates by longitudinally moving one plate with respect to the other.

The third possibility is to keep the area of the plates and the spacing fixed and to utilize the change in dielectric constant. A typical application is the liquid level gage shown in Figure 19. The dielectric constant of a liquid is significantly greater than that of air. As the liquid rises between the plates the capacitance increases very significantly. This basic technique can be used to measure the level of most liquids as well as confined gases. The capacitance change with level is essentially linear but the dielectric constant will vary with thermal changes. When the fluid is an electrolyte the surface of the liquid forms one plate and a second plate is suspended above it. As the liquid moves upward, the capacitance increases, but in a nonlinear manner.

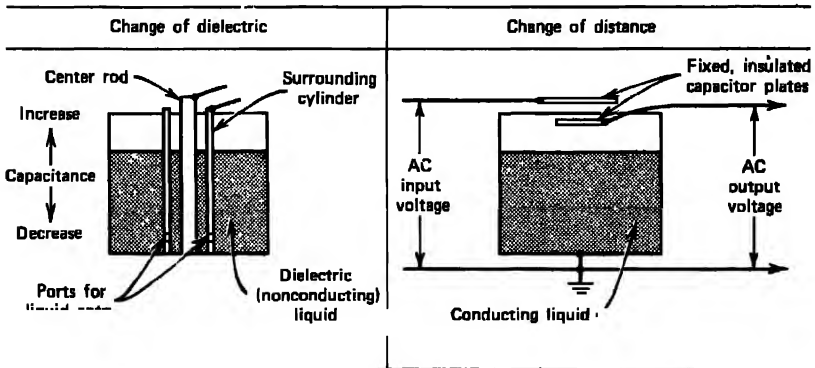


Figure 19. Liquid level gages. For nonconducting liquid use the transducer at left. It operates on change of dielectric level between center rod and surrounding cylinder. The transducer at right is for a conducting liquid. Change of distance between lower capacitor plate and surface of liquid changes capacitance. Varying airgap is the dielectric. (Courtesy of McGraw-Hill Book Company. From Reference 10.)

There are two methods generally used for measuring the output of a capacitance transducer. The first is to use the variations in capacitance to amplitude-modulate a carrier. The output is then amplified and demodulated in the conventional way. The second method is to use the transducer as part of a tuned circuit to vary the carrier frequency. The output is then detected using frequency modulation techniques.

The frequency response of a capacitive transducer can be computed as follows:

$$f_1 = \frac{1}{2\pi(C_T + C_W)R} \quad (9)$$

where f_1 = cutoff frequency (hertz)

C_T = capacitance of the transducer

C_W = capacitance of the wiring and all other parasitic effects

R = input impedance of the measuring circuit

The principal advantages of capacitive transducers are extreme sensitivity, excellent frequency response, and high output impedance. The limitations are the errors introduced by humidity, dirt accumulations between plates, spurious signals that are picked up by the leads, temperature drift, and relatively complex construction.

2.4. OPTICAL TRANSDUCERS

Optical measuring systems have been used in the laboratory since the time of Galileo. Modern technology has concentrated on repackaging them into sizes that are practical for industrial applications.

Most optical transducers are fundamentally photoelectric devices. The essential components are a light source, a detector, and an amplifier. Distance is determined by measuring the variation in light intensity from the source to the detector. There are three primary ways in which they are used (Figure 20):

1. As a go-no-go gage where the source of light is either completely blocked or completely transmitted.
2. As a proportional device to determine the transmissibility of various transparent or translucent materials. A photographic densitometer used to control exposure time is an example of this type of usage.
3. As a comparator to measure the deviation of light intensity or color from a standard reference. Color meters used for matching fabrics and paints employ this principle.

2.4.1. Light Sources

Three types of light sources are available: monochromatic, heterochromatic, and polychromatic (Figure 21).

Monochromatic radiation is concentrated at a given frequency or a narrow band of frequencies. Sodium vapor lamps and gallium arsenide diodes are examples.

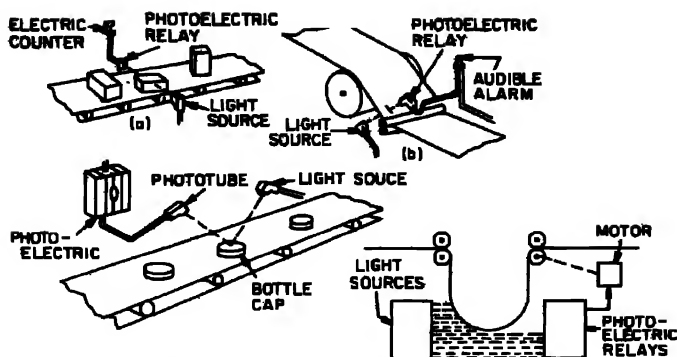


Figure 20. Applications for photoelectric devices. Typical examples of photoconductors as limit switches are counting moving objects on a conveyor (a) and detecting a break in a web (b). A simple inspection job that you can do with photocells is sensing for the presence or absence of a paper liner in a bottle cap. When the liners are present, the beam is not reflected. When a liner is missing, the reflected light causes a relay to operate and reject the cap. Maintaining proper loop depth in a moving web is a common control application of photoconductive cells. The amount of light striking the photocells creates a proportional current used to increase or decrease the motor speed of the web drive. (Courtesy of General Electric Company.)

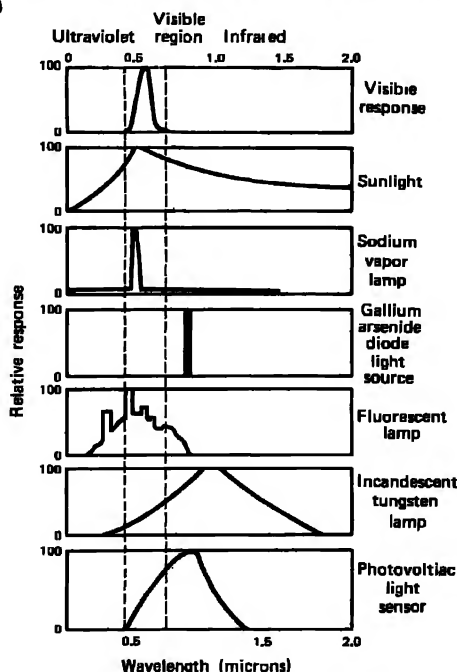


Figure 21. Comparison of light sources and photovoltaic light sensors. (Courtesy Hoffman Electronics. From Reference 11.)

Heterochromatic radiation is composed of several distinct bands, each at a characteristic frequency. This principle is used in emission spectrum work for identifying substances such as mercury, lithium, and hydrogen.

Polychromatic radiation covers a wide range of frequencies with no distinct bands.

System considerations largely determine the selection of the light source. Any installation requiring surveillance by an invisible source requires a gallium arsenide unit. If response of the human eye is a factor, sodium vapor or fluorescent sources would be the choice.

Photoelectric devices are often used by bounding light rays off the surface being monitored; the surface has a spectral sensitivity that makes it transparent to light of one frequency and reflective to other frequencies. If the instrument utilizes reflection principles, it is necessary to select a light source that will not be transmitted by the target.

By comparing the response curves of fluorescent and incandescent lamps we can understand why the former is more efficient for visible response.

In most cases the light source also contains a lens system. The optical system makes maximum use of the radiation produced by the light source. Typical systems employ one lens to form a series of parallel rays and a second lens to concentrate them on the receiver.

2.4.2. Detectors

The detector, or receiver, used in photoelectric systems is normally a photoconductive or photovoltaic cell. Photovoltaic cells are small power sources that generate a voltage when exposed to light. Photoconductive cells are devices whose conductance varies with the amount of light radiation falling upon them; they are similar to light-sensitive resistors. Usually photoconductive cells are 1000 times more sensitive than photovoltaic units, but they do require a power source.

When choosing a detector for a photoelectric system the following parameters should be checked:

1. Spectral response—the detector sensitivity should match the source as closely as possible.
2. Response time—the turn-on and turn-off times of sources and detectors should be compatible.
3. Life—whenever possible use hermetically sealed units for maximum reliability.

2.4.3. Amplifiers

Normally amplifiers are the least of the problems associated with optical transducers. The biggest problem is drift. The frequency response must be flat from DC to several kilohertz. Usually subminiature parts or integrated circuitry are specified. Heat sinking is also important.

2.5. FIBER OPTICS

One of the characteristics of optical transducers is that the source and detector must be aligned. In many instances the geometry of the problem makes this impossible. Fiber optics eliminate this requirement (Figure 22). They consist of thin strands of fiber bundles that are capable of transmitting light in much the same way as a pipe conducts water. See Figure 23 for operating principles.

A fiber optics transducer consists of a light source, a detector, and a bifurcated fiber optics cable. One half of the cable contains elements that transmit light from the source; the other half detects light that has been reflected off a given surface and "pipes" it back to the detector. The amplified detector output is displayed on a conventional or digital voltmeter. The system can be used to measure changes in distance as small as 5μ in. Frequency response is flat to 15 Hz.

The advantages of fiber optics are as follows:

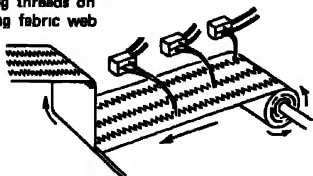
1. Ability to measure in confined spaces.
2. Transmission of light without appreciable heat.
3. One light source that can serve many measurements.
4. A diameter that, at present, is as small as $\frac{1}{8}$ in. Some of the limitations are as follows:

1. Fiber optics are composed of thin filaments, as small as 0.001 in., and may break when the bundle is flexed too often. This produces black spots in the bundle which transmit no light and decrease the linearity of the system.

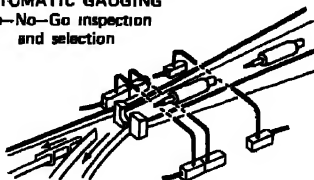
2. Since the light transmitted by fiber optics is less than that produced by other photoelectric devices, the amplifier gain must be higher, making parasitic effects such as capacitance, pickup, and thermal drift a more formidable problem.

3. If the measurement technique involves covering and uncovering the light source, the motion cannot produce a linear response, since the fiber optic cross-section is circular (Figure 24).

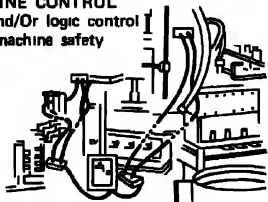
INSPECTION
Missing threads on
moving fabric web



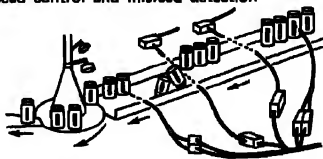
AUTOMATIC GAUGING
Go-No-Go inspection
and selection



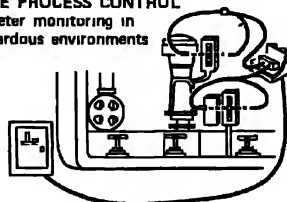
MACHINE CONTROL
Simple And/Or logic control
for machine safety



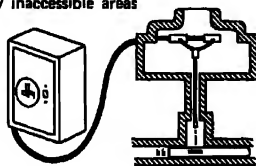
CONVEYOR CONTROL
Jam detection, counting, conveyor
speed control and misfeed detection



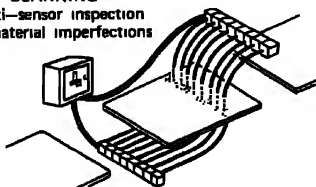
REMOTE PROCESS CONTROL
Meter monitoring in
hazardous environments



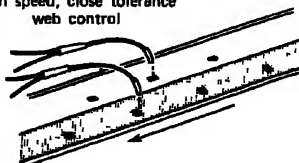
MONITORING
Internal shaft rotation in
normally inaccessible areas



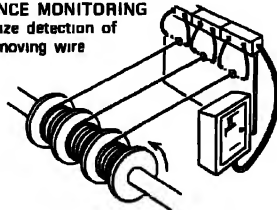
SCANNING
Multi-sensor inspection
for material imperfections



REGISTRATION CONTROL
High speed, close tolerance
web control



TOLERANCE MONITORING
Oversize detection of
moving wire



DIE PROTECTION
Progressive cut and carry
operation

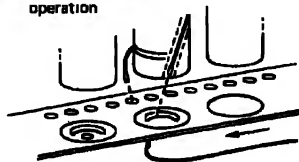
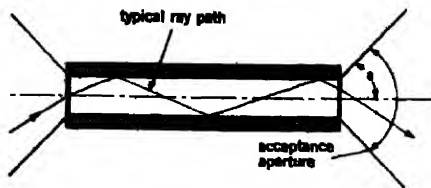


Figure 22. Applications for fiber optics. (Courtesy of Dolan-Jenner Industries, Inc. From Reference 11.)



$$NA = (n_{dc}^2 - n_{ds}^2)^{1/2}$$

or

$$NA = \sin \theta \text{ where:}$$

NA = numerical aperture

n_{dc} = INDEX OF REFRACTION OF CORE

n_{ds} = INDEX OF REFRACTION OF SHEATH

θ = HALF ANGLE OF ACCEPTANCE APERTURE



Figure 23. Fiber optics principles. A ray of light in a medium of high refractive index directed toward one of a lower index is refracted as it passes into the second medium. As the ray is inclined from the normal, it reaches a point in which it does not pass into the second medium but is totally reflected at the interface; it continues to be reflected until it leaves the end of the element. If the angle of the ray entering is greater than the maximum angle determined by the refractive indices of the core and surrounding sheath, the ray will not be totally reflected but will pass through the interface and be lost. The maximum angle, called the numerical aperture [NA], from the core axis for which total internal reflection occurs, indicates the light gathering capability of the system. (Courtesy of Bourns, Inc. From Reference 11.)



Figure 24. Fiber optics hardware. (a) Supply multiple illumination from a single light source, with multibranch light guides. (b) Provide input-output geometry conversions such as round-to-square and round-to-slit. (Courtesy of American Optical Company.)

2.6. INTERFEROMETERS

One of the most accurate techniques developed for linear measurements utilizes the interference of light. Until recently, interferometers have largely been laboratory instruments. Today, however, three basic types are available commercially: the laser, moiré, and microwave types. The oldest of the three is the laser interferometer, developed to check dimensions as small as several microinches.

1. *Laser.* To understand the principle of operation of this interferometer, consider the basic Michelson interferometer which consists of two mirrors, a beam splitter, telescope, path length adjuster, and a source of light (Figure 25). Monochromatic light from the source is formed into a parallel beam by a lens and projected toward the beam splitter. The splitter divides the ray in two, one directed at mirror *A* and the other at *B*. They advance toward mirrors *A* and *B* respectively, are reflected back to the beam splitter and then to the telescope. Note that the ray directed toward mirror *B* passes through the beam splitter only once, while the other ray passes through it three times. Plate *C* adjusts the two path lengths so that they can be easily sighted and "telescoped." Since the velocity of light is less in the beam splitter than in air, the two rays will be out of phase with respect to each other as they enter the telescope. If the rays, or waves, are in phase, they will appear as bright bands in the telescope; conversely, if they are out of phase, a dark band appears. By properly varying the geometry and distances of the mirrors, you can easily detect a change in distance of one fourth of a wavelength.

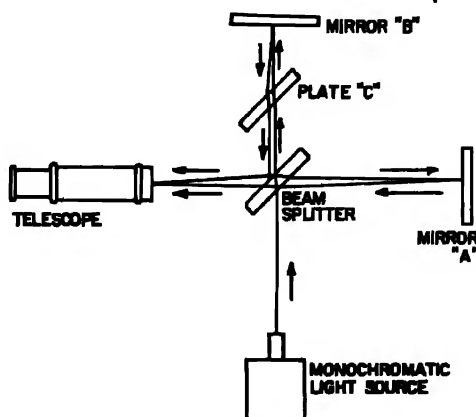


Figure 25. Michelson interferometer. (Courtesy of Benwill Publishing Company. From Reference 11.)

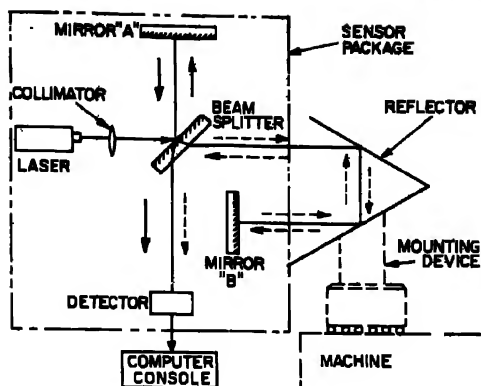


Figure 26. Laser interferometer on a machine tool. (Courtesy of Benwill Publishing Company. From Reference 11.)

Figure 26 shows a laser interferometer on a machine tool. The laser beam is directed through a collimator to the beam splitter, dividing the ray; one ray toward mirror *A*, where it is reflected back through the beam splitter that finally reflects it toward the photodetectors; the other ray toward a reflector prism where its path is twice reflected through 90° so that it strikes mirror *B*. Mirror *B* reflects the ray back to the reflector and then to the splitter, which finally reflects it toward the photodetectors. Path lengths of the two rays are therefore greatly different, and the sensors will detect a series of bands. Commercial instruments contain all elements, except the reflector, in one package. The reflector mounts on the object; for example, the bed of a milling machine. For each one eighth of a wavelength the bed moves, forming a band. The photodetectors sense the change and transmit the data to a counting circuit, which displays the change in position in millionths of an inch or in some other convenient units. The logic circuitry used can add or subtract so that motion in two directions is acceptable for the system without any backlash problems. The object being measured can move at a rate of 2 to 4 in./sec without any control complications.

2. *Moiré*. Lord Rayleigh originally suggested moiré fringe techniques for testing diffraction gratings. During the past few years, industry has finally perfected the apparatus to yield:

- (a) Location of the test surface with respect to some reference surface.
- (b) Angular position of the test surface with respect to the reference.
- (c) Radius of curvature of the test surface as a measure of deformation.

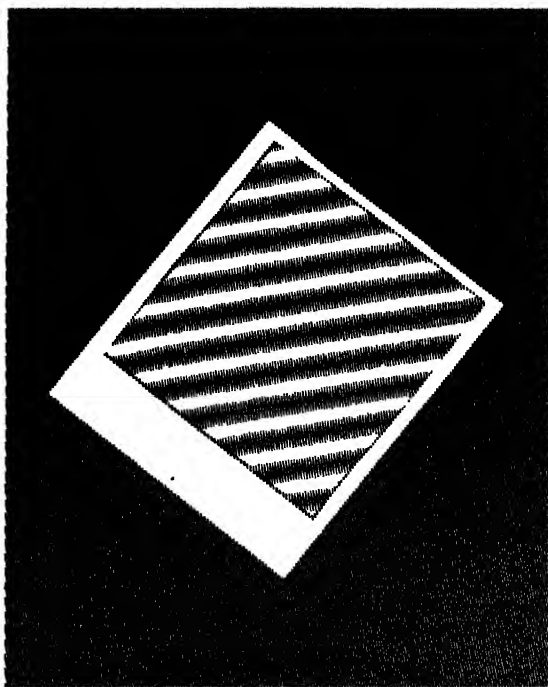


Figure 27. Moiré principle. (Courtesy of Perkin-Elmer Corporation.)

Proper use of these three properties makes it possible to map the contour of a surface within microinches (Figure 27). Moiré fringes are formed when the image of one grating falls on the image of a second grating, displaced by some small angle of rotation. Both gratings must contain the same number of bars per unit length. The fringes form at the intersection of the grating lines. Since the intersections are continuous across the gratings, the eye sees these intersections as a fringe. The spacing depends on the angle of intersection and the number of bars per unit length. Under the special condition that the angle of intersection is zero, no fringes form and the viewer sees a completely dark, a completely light, or a uniformly gray field, depending on whether the grating lines mesh, superimpose, or are in an intermediate position. As the angle between gratings increases from zero to about 3° , fringes appear, and the spacing between fringes decreases as the angle increases.

Figure 28 shows the Moiré interferometer. The laser beam is directed through a collimating lens to the polarizer grating with the shadow of the

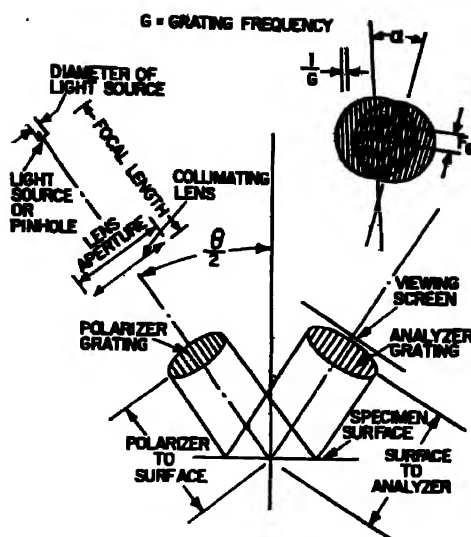


Figure 28. Moiré interferometer. (Courtesy of Benwill Publishing Company. From Reference 11.)

grating reflected off the test surface onto the analyzer grating to form the moiré fringes. Each of the gratings mounts in a holder which allows it to rotate about the interferometer's optical axis. The fringes are viewed on a screen above the analyzer grating with provision made for recording the pattern. Typical gratings used have 50 to 200 lines per inch.

3. *Microwave.* This instrument can be set up relatively quickly without too much attention to mechanical alignment of components. It consists of five components: a radio frequency source, phase comparator, transmitting and receiving antenna, reference path, and a micrometer-driven phase shifter. By observing the output of this instrument on a scope and inserting bandpass filters, the peak-to-peak displacement of each harmonic in a complex wave form can be analyzed; this is very useful in stress analysis work. This instrument is particularly suited for applications requiring high frequency response, since it operates on a carrier frequency of 35×10^9 Hz.

The principle of operation is shown in Figure 29. The RF source generates a carrier frequency of 35 GHz (10^9), modulated by a 100-KHz sawtooth generator. The signal is directed simultaneously in two parallel paths: one to a radiating and receiving antenna where the wave bounces off a target and is then received; the second to a reference channel containing a

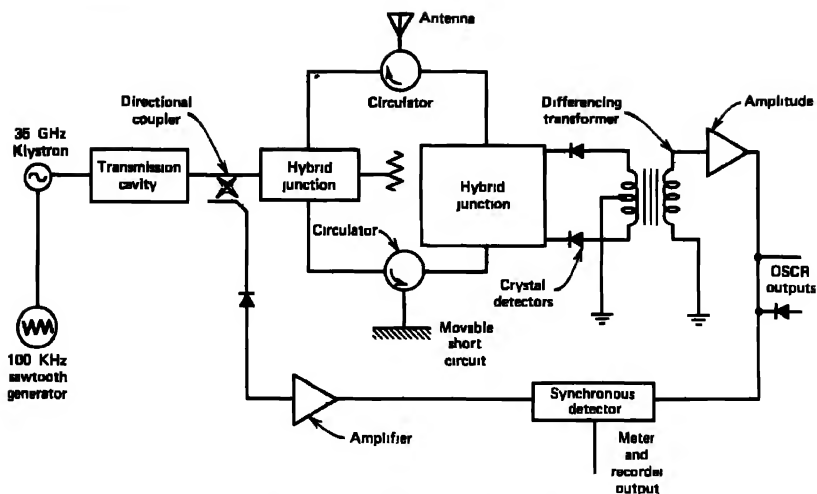


Figure 29. Microwave interferometer. (Courtesy of Weinschel Engineering Company.)

phase shifter and phase comparator. Since the two path lengths are unequal, a phase difference exists when the two signals are compared at the phase comparator, proportional to the distance from the antenna to the object being measured. The phase shift is reduced to zero by a phase shifter which aligns the phase position of the reference signal with that of the incoming signal. Since the phase angle depends on the position of the target, it can be calibrated to read directly in distance from the antenna to the target, practically expressed in microinches. The distance from the antenna to the target is not limited to very small distances, since we are usually interested in measuring variations in a dimension rather than the dimension itself.

2.7. BETA GAGES

Beta gages are the most widely used devices of gaging thickness (Figure 30). These instruments meet the requirement for a relatively small instrument that can be incorporated in modern automatic equipment. Unlike the older X-ray measuring technique, beta gages form an integral part of process equipment. They do not require much power, liquid coolants, or a special gas supply. Applications include controlling the production of paper, rubber sheet, plastic film, foil, and basic metals. Materials usually absorb beta rays

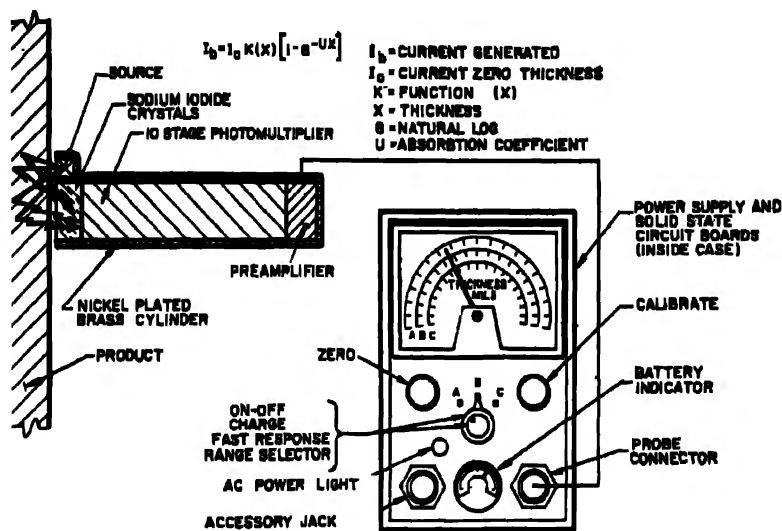


Figure 30. Beta gage. (Courtesy of Conrac Corporation. From Reference 11.)

in direct proportion to their specific weight. As a result, two types of measurements can be obtained: a readout on a weight basis where the thickness is constant and a readout on a thickness basis where the density is constant. This type of equipment typically produces measurements accurate within $\pm 1\%$. Drift is negligible.

Figure 30 shows that subatomic particles coming in contact with any material in their path may be absorbed, reflected back, or passed completely through the material. If a detector is placed on the same side of the material as the radiation source, it responds to the amount of particles reflected back; conversely, if it is on the opposite side of the material, it responds to the

amount of beta rays that pass through the material. For any given material, we can establish the degree of absorption, transmissibility or reflectivity during calibration procedures. The source of radiation is sodium iodide, krypton 85, or strontium 90. The detector is usually a probe containing a multistage photomultiplier or a semiconductor whose output is amplified and transmitted to a readout device or control console.

For a summary of linear sensor characteristics, see Figure 31.

2.8. ANGULAR DISPLACEMENT TRANSDUCERS

Although potentiometers, forms of angular inductors, and capacitive transducers are widely used as angular sensors in many servosystems today, the three primary tools used in precise systems are synchros, resolvers, and encoders.

2.8.1. Synchros

Synchro is a generic term originally used by the U.S. Navy to describe a rotary inductor in which variable coupling between the primary and the secondary is obtained by changing the relative orientation of the windings (Figure 32). This differs from the rotary variable differential transformer (RVDT) in which only the reluctance of the magnet circuit linking the windings is varied. The primary windings are wound on a rotor made of laminated magnetic stampings. The secondary windings are inserted in slots in the stator. Electrical connections between the rotor and stator are made through precision slip rings. Although the name "synchro" is universally used in the instrument field, trade names such as Selsyn, Microsyn, and Dielsyn are also commonly employed.

The classical synchro system consists of two units: a synchro transmitter and a synchro motor or receiver. The transmitter develops a signal as a result of the displacement of the rotor with respect to the stator. The synchro motor receives the signal and converts it into an output torque; it continues to rotate until its rotor is electrically aligned with the transmitter. Here is how this is accomplished (Figure 33). Initially the S_2 winding of the stator is positioned for maximum coupling with the primary of the rotor. Its voltage is V . The coupling between S_1 and S_3 of the stator and the primary is a cosine function; the effective voltages in these windings are proportional to the cosine of 60° , or $V/2$. So long as the rotors of the transmitter and receiver remain in this position, no current will flow between the windings because of the voltage balance. When the rotor of the transmitter is moved to a new position, the voltage balance is disturbed. Assume that the rotor of the

Type	Range	Measurement	Repeatability (inch)	Linearity	Materials	Comments
Proximity						
Inductive	0 - 6"	position	$\pm 0.001 - 0.005$	1 - 5 %	Ferrous	Good frequency response
Magnetic	0 - 2"	position	± 0.005	ON - OFF	Ferrous	Cable length not critical
Capacitive	0 - 12"	position	± 0.005	Hyperbolic	All	Low frequency response
Ultrasonic	0 - 20'	position	± 0.01	ON - OFF	All	Needs auxiliary equipment
Photoelectric	0 - 200'	position transmissibility comparator	$\pm 10^{-3} - 10^{-6}$	0.1 - 1 %	All	Requires moderately careful alignment
Fiber optics	0 - 1/4"	position	± 0.001	± 15 %	All	Excellent for confined areas
Interferometers						
Laser	0 - 20'	position	$\pm 10^{-4} - 10^{-6}$	0.01 - 0.1 %	All	A prism must be attached to object
Moire	0 - 1"	surface contour	$\pm 10^{-4} - 10^{-6}$	0.01 - 0.1 %	Reflective	Can be combined with high speed camera recording
Microwave	0 - 1"	position	$\pm 10^{-4} - 10^{-6}$	0.01 - 0.1 %	All	Excellent frequency response
Beta-Radiation	0 - 3"	thickness	± 0.0002	± 1 %	All	Designed for process control industry

Figure 31. Characteristics of sensors. (Courtesy of Benwill Publishing Company. From Reference 11.)

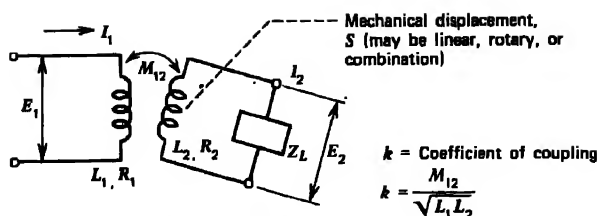
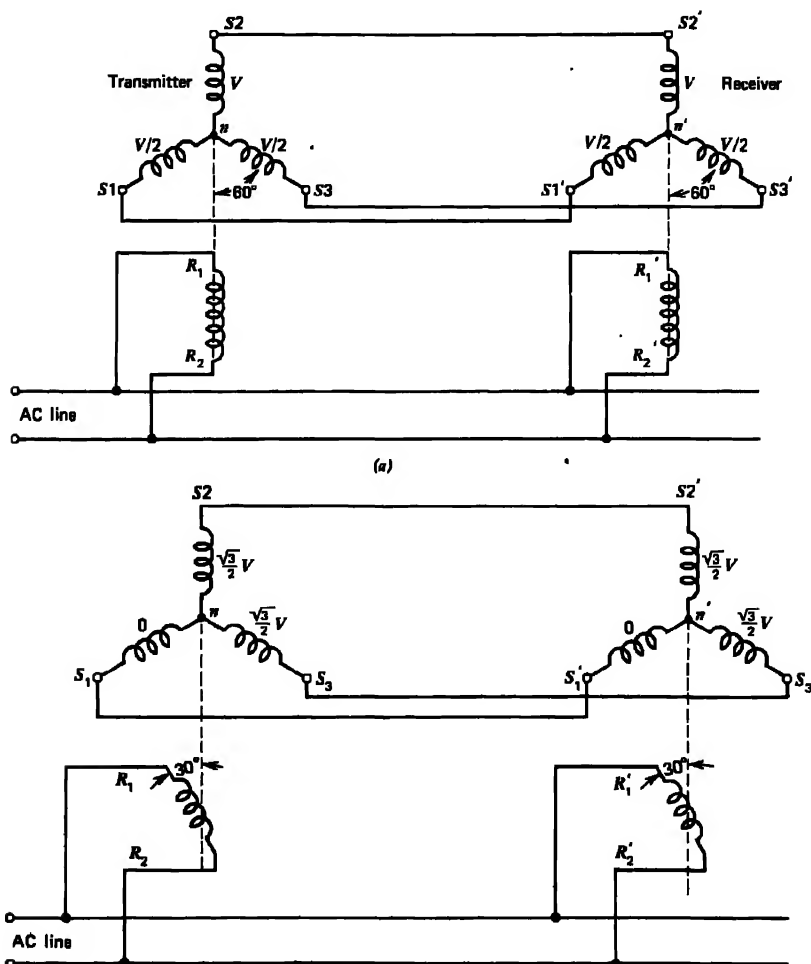


Figure 32. Schematic of a pair of mutually coupled coils. I_1 , L_1 , R_1 , E_1 = current, inductance, resistance, and voltage across the primary coil. I_2 , L_2 , R_2 , E_2 = current, inductance, resistance, and voltage across the secondary coil. M_{12} = mutual coupling between the two coils. (Courtesy of Benwill Publishing Company. From Reference 11.)



transmitter is turned 30° ; the stator winding voltages will be changed to 0, $\sqrt{3}/2$, and $\sqrt{3}/2$ V respectively. The voltage unbalance between the windings of the transmitter and receiver causes current to flow between the windings and creates a torque that tends to rotate the receiver rotor until it is once again electrically aligned with the transmitter.

There are two types of synchro systems: torque and control types. Torque-type systems are used only to drive very light loads, such as pointers. Paradoxically, torque-type systems have very little output torque. When large torques and high accuracies are required, control-type synchros are used (Figure 34).

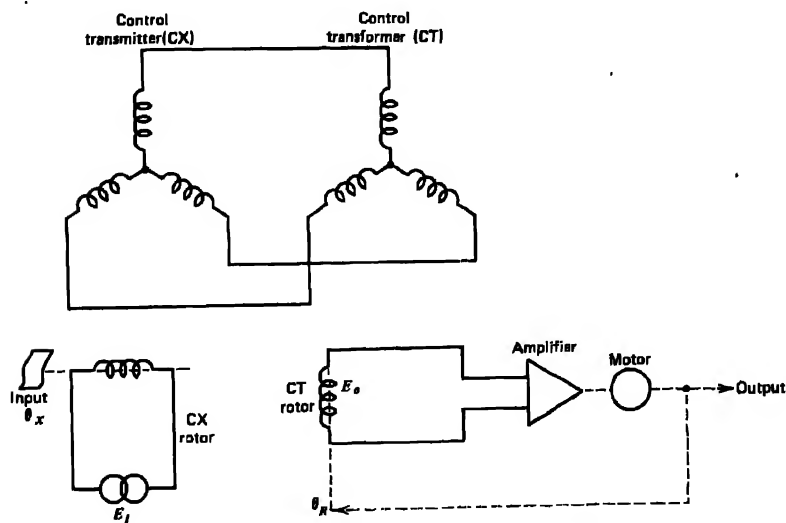


Figure 34. Typical control synchro system—single phase AC is used to excite the rotor of the CX. The three-phase voltage developed in the CX stator, which is a function of the CX rotor position, is transmitted to the CT stator. The voltage developed across the winding of the CT rotor is amplified and used to drive a servomotor. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

A simple torque-type system consists of a transmitter (TX) and a receiver (TR). When it is necessary to add or subtract a fixed angle between them, a differential synchro (TDX) is inserted in the circuit (Figure 35). A torque receiver shaft aligns itself with the position of the transmitter or differential synchro. Thereafter it behaves like a spring; when it is moved from its

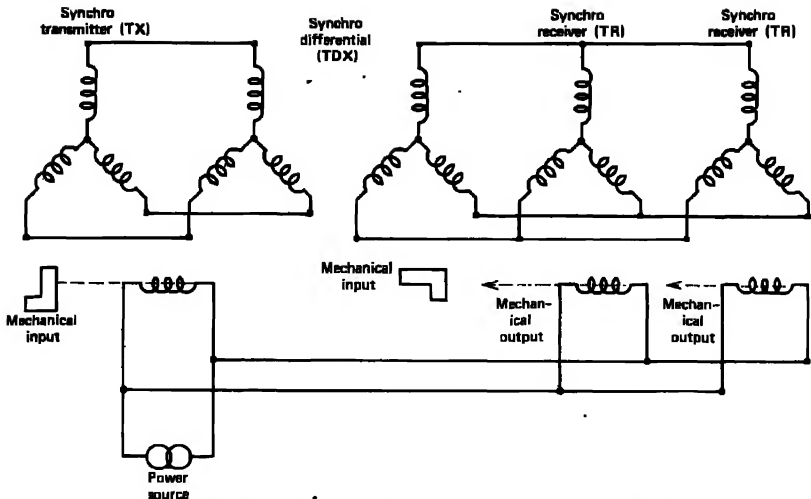


Figure 35. Torque synchro system. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

synchronous position it exerts a restoring torque proportional to the displacement. Bearing friction is primarily responsible for preventing a receiver shaft from snapping into exact synchronism.

The vast majority of synchros in use today are part of control-type systems that utilize a servoamplifier. The following components are used in this type of system (Figure 36):

1. *Control transmitter (CX)*. The rotor is mechanically positioned to transmit electrical information corresponding to the angular position of the rotor with respect to the stator. In terms of impedance, this component is constructed to operate with control transformers or control differentials.

2. *Receiver or servomotor* (not illustrated in Figure 36), sometimes called the synchro receiver (CR). This converts the electrical information supplied to it into a mechanical shaft output.

3. *Control differential transmitter (CDX)*. This device receives two inputs, one electrical and one mechanical. As the rotor is mechanically positioned with respect to the stator, the coupling between the two windings varies as the sine of the displacement angle between them, and the net output voltage is the difference of the two inputs. The CDX is useful in applying a fixed correction voltage to a control transmitter signal.

4. *Control transformer (CT)*. This instrument receives an input from a CX or CDX and converts it into an output voltage that is proportional to

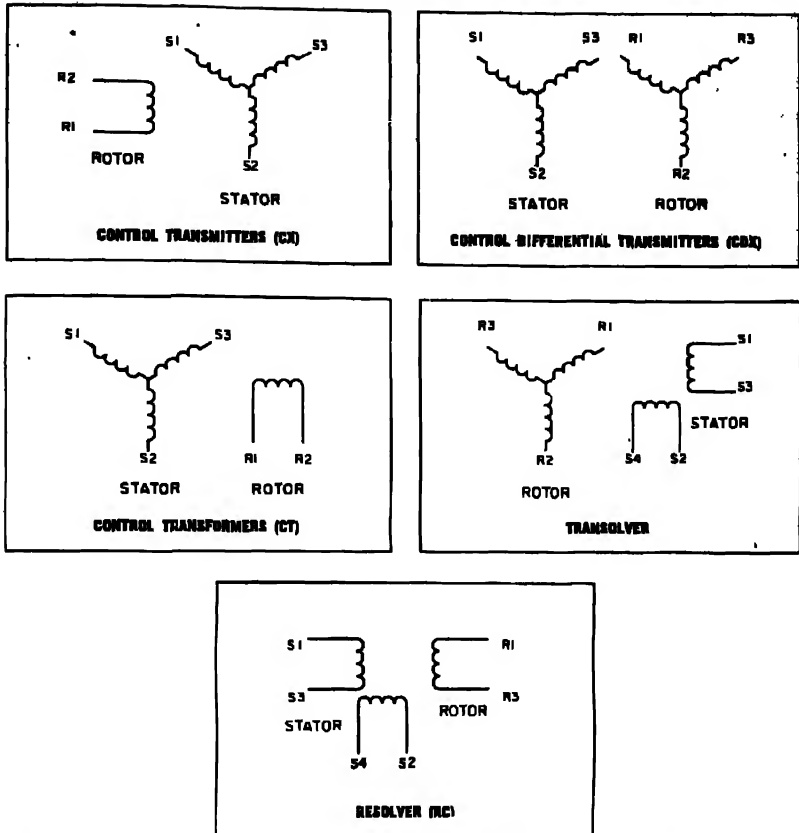


Figure 36. Synchro terminology. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

the sine of the angle between the rotor and stator. It is fundamentally different from a CDX, since the output of a control transformer is single-phase while that of the CDX is three-phase. Its prime use is as an error indicator. If the stators of a CX and CT are properly aligned, any difference in their respective rotor positions will result in an error signal from the CT rotor winding. By contrast, the CDX is used to modify the three-phase output of a CX.

5. **Resolver (RC).** A resolver is used to convert the angular position of a shaft into cartesian coordinates. The output of the instrument is in the format of two signals, one proportional to the sine and the other proportional to the cosine. See Section 2.8.2 for more details.

6. *Transolvers (CSD).* This hybrid device has a two-phase rotor and a three-phase stator. It can be used as a transmitter, control transformer, and a resolver. When one rotor phase is excited and the other shorted it can be used as a transmitter. For use as a CT, the stator is excited from another three-phase device (CX or CDX), and the output is obtained from one rotor phase. The rotor phases are 90 electrical degrees apart; consequently a sine and cosine output can also be obtained from a transolver.

7. *Servo amplifier.* The input impedance must be high with respect to the CT output impedance. The gain is determined by a servo analysis. Standard commercial units designed for these applications are readily available. A servoamplifier is composed of high-density potted, transistorized units. They are available with gains from unity to 10,000 and phase shifts of 0 to 90°. The 90°-phase shift is useful in eliminating the phase-shifting capacitor in the fixed phase of a servomotor. Typical characteristics are shown in Figure 37.

	C70 3189 001	C70 3105 Series
Input Power (watts) @ zero signal	1	1.3
@ rated output	7	7.7
Output Power (watts)	3.5	3.5
Input Signal (v rms) (max.)	30	26.5
Input Impedance (ohms)	10,000 min.	R_1 (external) + 500
Input Voltage (vdc) ($\pm 10\%$)	28	27.5 (25 for 013)
Rated Output Voltage (v rms)	36	40 (36 for 013)
Maximum Gain	5000 ± 1 dB	1000 $\pm 15\%$ *
Gain Stability (dB)	± 4 (-55°C to 0°C) ± 1 (0°C to $+125^\circ\text{C}$)	± 2
Output Impedance (ohms) (max.)	100	200
Phase Shift ($^\circ$) linear region	0 ± 10	0 ± 10
@ saturation	0 ± 15	0 ± 15
Operating Temperature Range ($^\circ\text{C}$)	-55 to $+125$	-55 to $+125$
Storage Temperature Range ($^\circ\text{C}$)	-65 to $+125$	-55 to $+125$

* $R_1 = 5000$ ohms

Figure 37. Typical servoamplifier characteristics. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

A typical control circuit appears in Figure 38. The command signal is mechanically introduced by rotating the rotor of the CX. The error signal developed in the CT is directed to the input of the amplifier, which in turn drives the servomotor; the servomotor is geared to a load and also to the rotor of the CT. The servomotor drives the load until the rotor of the CT is again in phase with command voltage generated by the CX. At this point there is no longer an output from the CT, and the motor is deenergized.

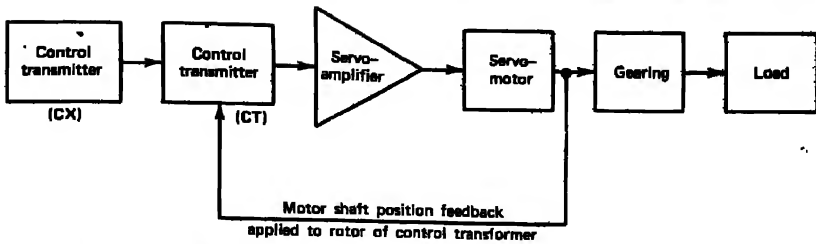


Figure 38. Typical servomechanism using synchro error detectors.

2.8.1.1. Selection of Synchros. The place to start is the system block diagram. First analyze the load to determine the inertia, frictional torque, and required acceleration characteristics. Next, select the servomotor and gear train. Consider the synchro and amplifier last, since they must be compatible with the rest of the system. Assume that the synchros in this example are a CX and CT. See Reference 17 for a detailed analysis of this work.

INPUT POWER. Most control transmitters operate on 28 V DC or 115 V AC with an excitation frequency of 60 or 400 Hz. Special units operate on odd voltages such as 6.3 or 90 V. Control transformers and differential transformers operate on the output voltage of CXs, which is typically 11.8 or 90 V.

SENSITIVITY. Sensitivity is the output voltage from the stator per degree of rotor rotation (mV/degree). Commercial units offer a selection of sensitivities from about 40 to 1600 mV/degree. Since the CX is the basic sensor, the overall response of the system is determined by the sensitivity selected. The sensitivity of the CT is equally important; it determines the response time of the synchro pair.

TRANSFORMATION RATIO (TR). Transformation ratio is the output voltage across two windings divided by the input voltage of a synchro at rated load at maximum coupling. This number is used in determining the output of the CX at maximum coupling. Optimum design considerations of the unit determine the TR which is generally less than unity for a CX and greater than unity for a CT. Practical tolerance on TR is about ± 1 to 4%.

INPUT IMPEDANCE. The input impedance of a CX should be selected for compatibility with the power supply available and for a minimum reactive component. The resistive component rarely presents a problem but the reactive component can cause trouble, since the greater the reactive component, the lower the efficiency. The input impedance of a CT or CDX must

be compatible with the output impedance of the CX. Normally CXs and CTs are selected so that one CX can easily drive two or three CTs.

Most synchro manufacturers publish application data sheets that list the overall response of a combination of CXs and CTs. Using these sheets saves considerable slide rule work and eliminates wasted time on mismatched or obsolete components.

OUTPUT IMPEDANCE. The output impedance of a CX is normally compatible with most CDXs and CTs of the same size and the same voltage rating. The output impedance of the CT must be compatible with the input impedance of the amplifier it drives. Normally, this presents no problem since servoamplifiers can be obtained with input impedances from 5 K to 1 M Ω .

PHASE SHIFT. The lower the phase shift, the better. Most good units have phase shifts of 5° or less.

NULL VOLTAGE. The lowest open circuit residual voltage measurable as the shaft rotates about its electrical zero position is comprised of two components: quadrature fundamental null and harmonic voltages. Quadrature fundamental null is residual voltage having the same frequency as the excitation and in-time phase quadrature with the output voltage at maximum coupling. The voltages contributed by higher harmonic waveforms comprise the other components of null. Predominantly third harmonic, they combine with the fundamental to yield total average null. The amount of null tolerable in a system depends on the amplifier gain and its ability to discriminate between quadrature null and the in-phase signal. Total null is normally permitted to be 50 to 60% greater than the maximum permissible fundamental null.

ELECTRICAL ZERO. Every rotor position has a corresponding electrical position. The mechanical rotor position minus the electrical position gives the error at a given rotor position. Output voltages from a synchro winding passes through two nulls when rotated through 360°, making it necessary to zero each synchro physically when aligning servosystems. When representing error curves graphically, electrical zero is the datum from which all error is plotted.

FRICTION. The old adage about getting exactly what you pay for is especially true for synchros. There are no cut-rate prices on precision bearings and they determine the friction level of a synchro. A good synchro will use a class (ABEC) 5 or 7 bearing. High bearing friction will have exactly the same effect as high null voltage; the error of the unit is increased. Typical friction torques for size 8 to 10 synchros are in the range of 3 to 4 g-cm.

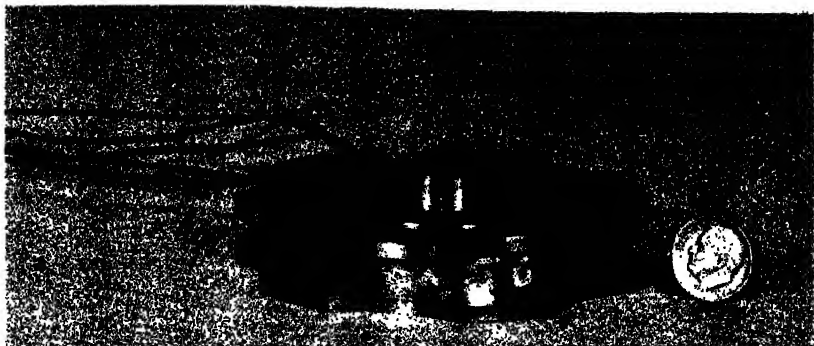


Figure 39. Pancake-type synchro. (Courtesy of Reeves Instrument Company. From Reference 15.)

MOMENT OF INERTIA. The rotor moment of inertia is largely a function of the ingenuity of the designer. The lower the inertia, the quicker the servo response. Typical values are about 1 to 2 g-cm² for small units.

RUN-OUT. Shaft eccentricity or run-out play havoc on systems where gearing is employed. Precise units hold run-out to 0.001 in. TIR. Shaft end play is about 0.0005 to 0.0015 in.

ENVELOPE NOMENCLATURE. Synchro diameters are normally expressed in tenths of an inch. For example, a size 22 unit has an outside diameter of about 2.2 in. The normal range is from size 5 to size 30.

CONFIGURATION. The normal configuration for a synchro is a cylinder where the length is approximately $1\frac{1}{2}$ times the diameter. However, for applications in inertial guidance platforms where the synchro must be attached to a relatively large shaft, the "pancake" design has been developed (Figure 39). It is available in aluminum, stainless steel, and beryllium copper housings.

2.8.2. Resolvers

Resolvers are used to convert the angular position of a shaft into cartesian coordinates. The output of the instrument is in the form of two signals, one proportional to the sine of the angle and the other proportional to the cosine. Resolvers are used for computing, phase shifting, position indication, and data conversion.

A resolver is a very precise electromagnetic device composed of two rotor windings and two stator windings within an accurately machined magnetic

structure. A second pair of stator windings is included in some units for feedback compensation. The rotor windings are electrically mutually perpendicular and the stator windings are wound in a similar manner. Accuracies in the order of 0.01% are typical for this instrument.

2.8.2.1. Basic Theory. When one stator winding is excited and the other is shorted, the following output is obtained at maximum coupling (Figure 40):

$$E_{R1-3} = E_{S1-3} \cos \theta \quad (10)$$

$$E_{R2-4} = -E_{S1-3} \sin \theta \quad (11)$$

or

$$E_{R4-2} = E_{S1-3} \sin \theta \quad (12)$$

When two stator windings are excited the outputs are as follows:

$$E_{R1-3} = E_{S1-3} \cos \theta + E_{S2-4} \sin \theta \quad (13)$$

$$E_{R2-4} = E_{S2-4} \cos \theta - E_{S1-3} \sin \theta \quad (14)$$

When two rotor windings are excited the results are:

$$E_{S1-3} = E_{R1-3} \cos \theta - E_{R2-4} \sin \theta \quad (15)$$

$$E_{S2-4} = E_{R2-4} \cos \theta + E_{R1-3} \sin \theta \quad (16)$$

For all equations θ = the angular displacement of the rotor.

The construction of a resolver is similar to a two-phase, two-pole, wound rotor induction motor. Stator windings are supplied with an alternating

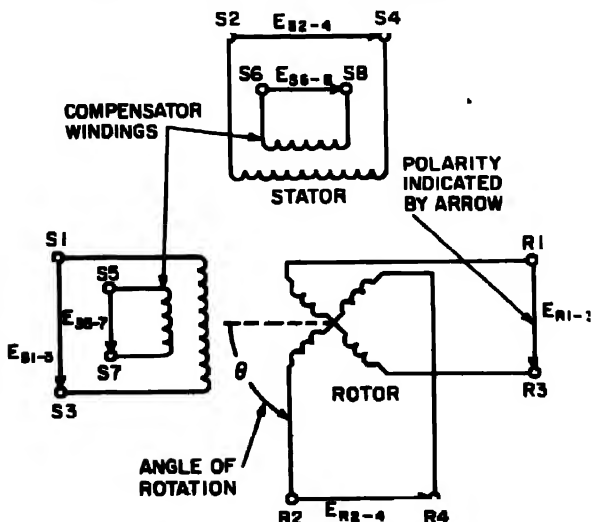


Figure 40. Standard schematic for four- and six-winding (compensated) resolvers. (Courtesy of Sperry Rand Corporation.)

voltage that produces a magnetic flux which induces voltages in the two rotor windings. The output voltage is proportional to the stator voltage and the coupling between rotor and stator. Because of the way in which the coils are wound, the rotor winding voltages are proportional to the sine and cosine of the rotor angle.

Resolvers are classified in two groups: computing resolvers and synchro resolvers. Computing resolvers are used for generating sine, cosine, and tangent functions as well as for solving geometric relationships such as right triangles (Figure 41). Synchro resolvers are used for data transmission; they

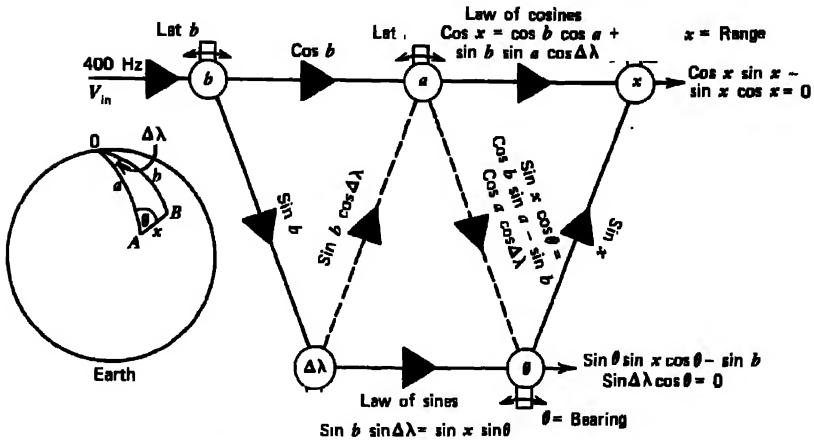


Figure 41. Use of a resolver to solve a problem in spherical trigonometry. *Given:* point A with its latitude and longitude; point B with its latitude and longitude; $\Delta\lambda$ = longitude of point A minus the longitude of point B; θ = bearing angle = $\angle OAB$. Find the range x = distance from point A to point B. Note that a is found by subtracting the latitude of point A from 90° and b is found by subtracting the latitude of point B from 90° .

Method: Find the length of two sides of a spherical triangle (a and b) and the included angle ($\Delta\lambda$) and solve for the third side (x). Solve for x so that the mechanical position of the rotor of the last resolver is proportional to the range.

Problem: Given two objects at latitude a and b , find the range, x , between them; θ = bearing. At resolver b , the input is latitude b , and the outputs are $\sin b$ and $\cos b$. At resolver $\Delta\lambda$, where $\Delta\lambda$ is the difference between latitudes a and b , the inputs are $\Delta\lambda$ and $\sin b$. The outputs are (1) $\sin b \cos \Delta\lambda$ and (2) $\sin b \sin \Delta\lambda = \sin x \sin \theta$ (law of sines). At resolver a , the inputs are latitude a , $\cos b$, $\sin b \cos \Delta\lambda$. The outputs are (1) $\cos b \cos a + \sin b \cos \Delta\lambda \sin a = \cos x$ (law of cosines) and (2) $\cos b \sin a - \sin b \cos \Delta\lambda \cos a = \sin x \cos \theta$ (trigonometric identity). At resolver θ , the inputs are $\sin x \cos \theta$ and $\sin x \sin \theta$. The outputs are (1) $\sin x \cos^2 \theta + \sin x \sin^2 \theta = \sin x$ (identity) and (2) $\sin \theta \sin x \cos \theta - \sin b \sin \Delta\lambda \cos \theta = 0$. At resolver x , the inputs are $\cos x$, $\sin x$. The outputs are (1) x , in the form of a shaft position, and (2) $\cos x \sin x - \sin x \cos x = 0$. The shaft rotation is caused by an auxiliary servo that drives the rotor to the position corresponding to the equation $\cos x \sin x - \sin x \cos x = 0$. A similar servo is used with the θ resolver. (Courtesy of Clifton Precision Products, Division of Litton Industries. From Reference 16.)

perform the same functions as synchro CXs, DXs, or CTs, but with better accuracy.

A computing resolver is one designed specifically for high-accuracy computations. It contains compensation to correct for variations in transformation ratio and phase shift and for the results of temperature, voltage, and frequency variations. An extra winding is provided to supply negative feedback in the circuit.

Synchro resolvers are sometimes called four-wire or two-phase synchros to emphasize the fact that these units are intended for use similar to that of a synchro. The specific types of synchro resolvers are the resolver-transmitter (RX), resolver differential (RD), and resolver transformer (RC). The RX uses a single-phase primary, although it is almost always supplied with two primary windings. The second winding is short-circuited to provide better accuracy and lower null voltages. The RC usually uses only one of the two secondaries supplied with the unit. When used in a null seeking servoloop, the spare winding is connected to a load identical to that used in the other secondary. Resolvers are easier to manufacture than synchros when the required accuracies are better than a few minutes of arc. Synchro resolvers are available in sizes 8 through 11 with accuracies of 10 sec or better.

The guidelines previously given for selecting synchros are also applicable to resolvers. In addition, resolvers have some special "bugs" of their own. The output voltage of a resolver is not always a perfectly sinusoidal function. This is due to variations in coupling between the windings and to the imperfect distribution of the coils. The primary winding must develop a flux density distribution in the air gap between the rotor and stator as close as possible to a sinusoid. The principal limitation is the finite number of slots which lead to a steplike approximation of a sinusoidal flux distribution. By matching primary and secondary distributions, primary harmonics may be canceled or minimized.

Transformation ratio and phase shift errors are also problems. Transformation ratio and phase shift taken together constitute a complex transformation ratio. In normal production runs neither the transformation ratio nor the phase shift is precisely controlled. Subsequent selection procedures may be used to obtain resolvers where these parameters fall within a narrow tolerance. The most important criterion for production units is that the transformation ratio and phase shift of each secondary be as close as possible.

Misalignment of the electrical zero axis of the primary and secondary windings is important and difficult to eliminate. To minimize this effect, maximum uniformity of the magnetic circuit is required. The rotor and stator must be mechanically rigid and remain undisturbed by assembly

operations as well as stress resulting from shock, vibration, and temperature excursions.

2.8.2.2. Resolver Applications. VECTOR RESOLUTION. The resolution of a vector into its components is shown in Figure 42. One stator winding is excited by a voltage proportional to the vector magnitude; the rotor is mechanically turned from electrical zero by an amount equal to the vector angle θ . This produces two output voltages, one from each secondary, that are proportional to the orthogonal components. Equations (10) and (11) describe the results.

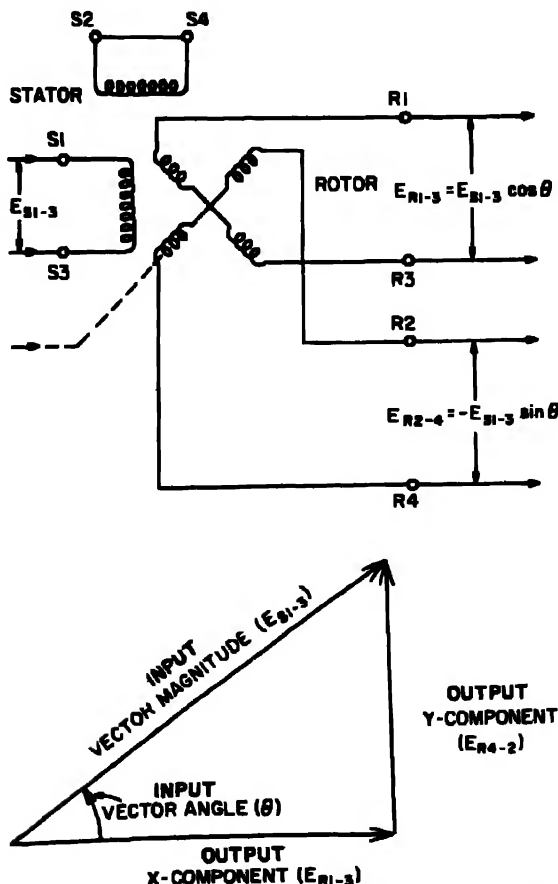


Figure 42. Vector resolution. (Courtesy of Ford Instrument Division, Sperry Rand Corporation.)

VECTOR COMPOSITION. The previously described operation can be inverted by connecting one secondary winding to a servoloop (Figure 43). The output winding is used as a null indicator in series with a servo-amplifier. The rotor is mechanically geared to the servomotor. The two components of the vector are fed to the two primary (stator) windings and the rotor output is described by equations (13) and (14). The output voltage from terminals $R2-R4$ energizes the amplifier and causes the motor to drive the rotor of the resolver until the null position is reached. The output from terminals $R1-R3$ is then analogous to the magnitude of the vector; the angle of the rotor displacement is equal to the associated vector angle. The same servosystem can be used to transform rectangular coordinates into polar coordinates, compute the arc tangent and hypotenuse of a right triangle, or add two perpendicular vectors.

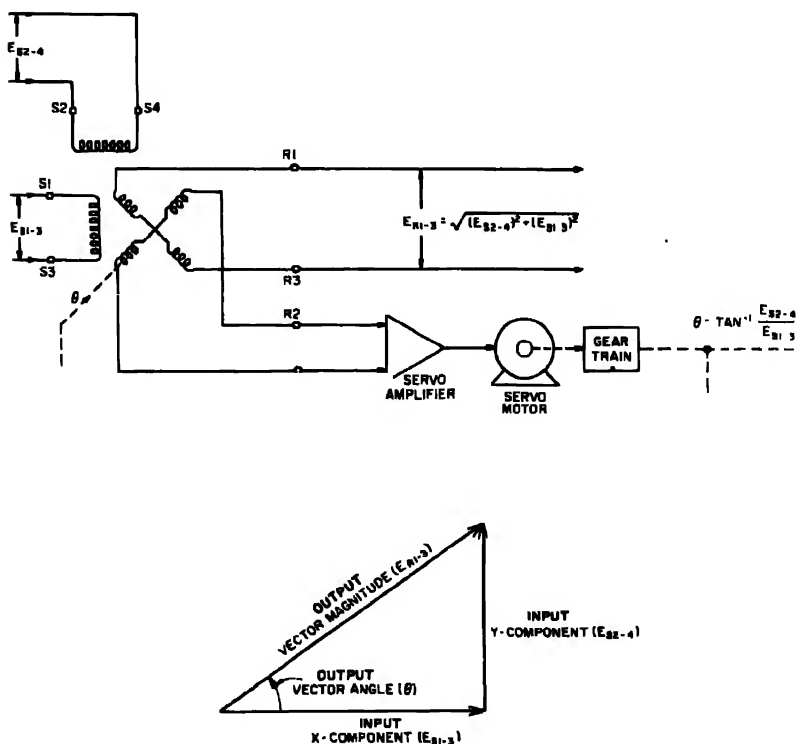


Figure 43. Vector composition. (Courtesy of Ford Instrument Division, Sperry Rand Corporation.)

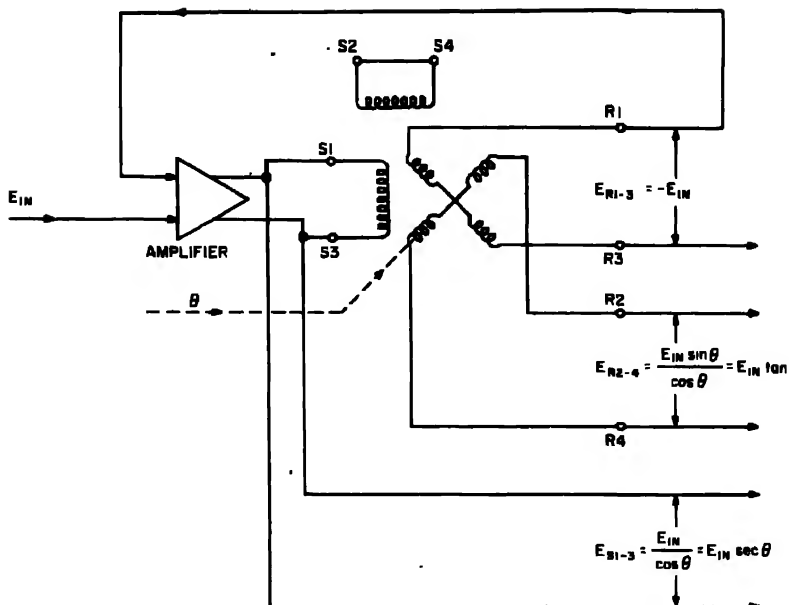


Figure 45. Secant and tangent computation (Courtesy of Ford Instrument Division, Sperry Rand Corporation.)

obtained from more complex systems. Special units are required that are precisely trimmed, and the associated power supply must be well regulated. The resolver is usually equipped with compensator networks to minimize temperature effects. The networks include thermistors, capacitors, inductors, and resistors to negate the effects of temperature and load variations. As many as five resolvers can be cascaded in this manner. Figure 46 shows such a system. Coordinates X , Y , and Z , measured in the reference frame of a moving vehicle, are translated to earth-stable coordinates by rotating the reference frame through angles equal in magnitude to the roll, pitch, and yaw of the vehicle.

PULSE AMPLITUDE CONTROL AND PULSE RESOLUTION. A resolver can resolve pulses into components proportional to the sine and cosine of the rotor angle, controlling pulse amplitude without substantially distorting the shape. A compensated resolver has a frequency response of at least 100 kHz; consequently, pulses as short as 10 μsec can be handled. The circuit shown in Figure 47 is used in radar, sonar, communications, navigation, and other

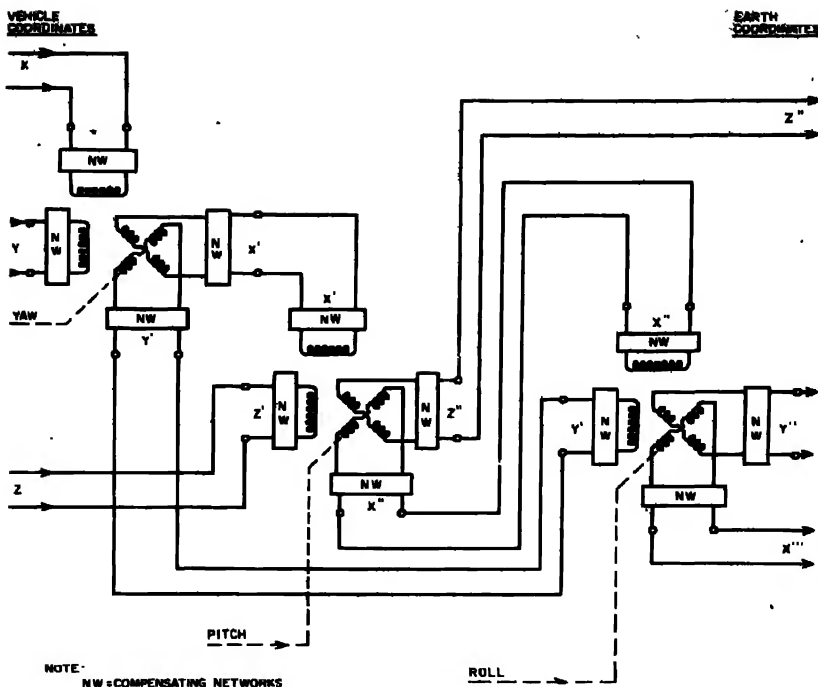


Figure 46. Cascaded amplifierless resolver system. (Courtesy of Ford Instrument Division, Sperry Rand Corporation.)

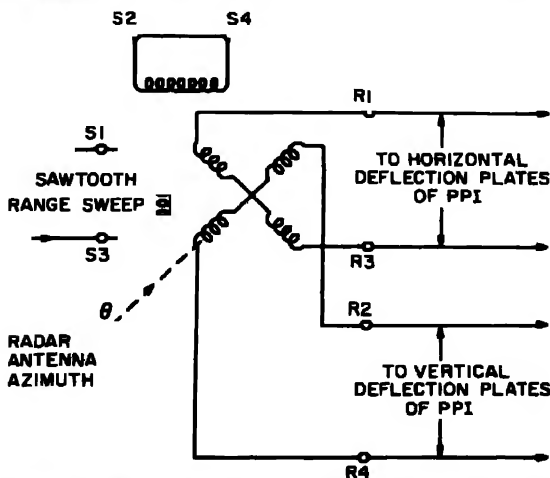


Figure 47. Radar range sweep resolution. (Courtesy of Ford Instrument Division, Sperry Rand Corporation.)

pulsing systems. The sawtooth range sweep pulses generated in the radar set are converted by the resolver into suitable deflection voltages for the position indicator. When isolation amplifiers and damping resistors are used, a sweep linearity within 0.1% can be achieved.

PHASE SHIFTERS. Typical applications of the resolver as a phase shifter are shown in Figures 48 and 49. In Figure 48 two voltages of equal amplitude but differing 90° in phase are applied to the stator windings. The rotor windings supply outputs that are offset from the inputs by an angle equal to the rotor angle θ . The phase shifter shown in Figure 49 operates from a single-phase source having noncritical voltage regulation. A double resistance-capacitance network is connected across the rotor windings. Values of R and C are selected to obtain a high impedance across each rotor winding and to satisfy the relationship

$$R = \frac{1}{2\pi fC}$$

where f = excitation frequency (hertz).

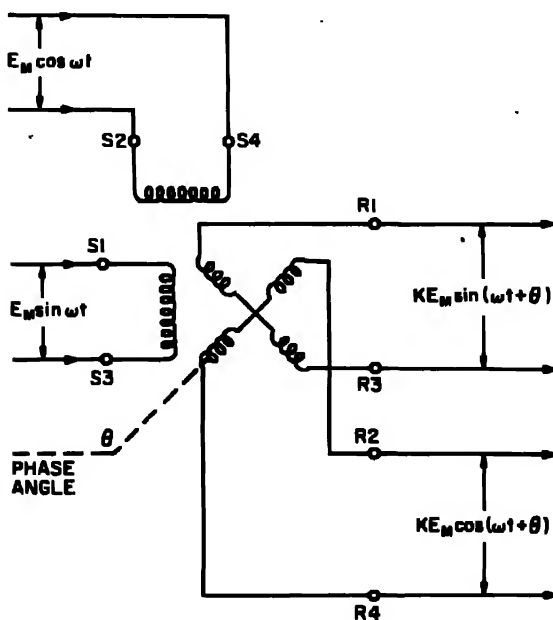


Figure 48. Phase shifter utilizing two-phase excitation. (Courtesy of Ford Instrument Division, Sperry Rand Corporation.)

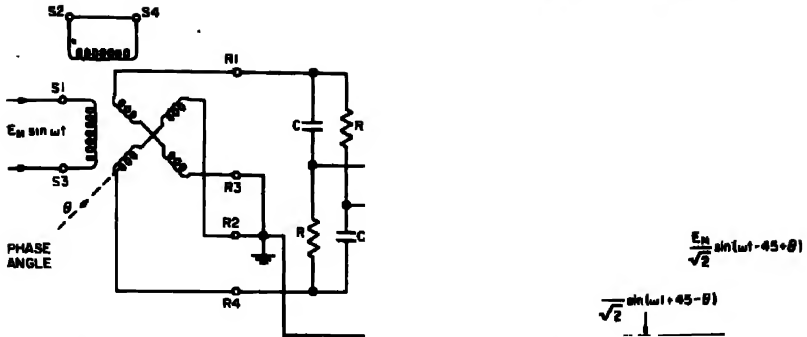


Figure 49. High-accuracy phase shifter with single-phase excitation. (Courtesy of Ford Instrument Division, Sperry Rand Corporation.)

The expressions shown in Figures 48 and 49 for the output voltages are based on maximum coupling. Under these conditions θ is measured from electrical zero. The constant offset of 45° can be removed by an opposite offset to the resolver rotor.

DIGITAL APPLICATIONS. For high-accuracy angular measurement and control the circuit shown in Figure 50 can be used to produce a digital output that accurately represents the rotor angle. Clock pulses are gated in agreement with the phase difference between the outputs of the RC networks to obtain a series of pulses that correspond in number to the rotor angle. This arrangement is useful in electric dividing heads where accurate angular measurement is required. The digital output can be transmitted without loss in accuracy for remote indication or control.

2.8.3. Encoders

The third basic angular transducer, the encoder, is covered in Chapter 3.

2.9. VELOCITY TRANSDUCERS

The most common application for velocity transducers is on vibration tables where an object is subjected to sustained harmonic motion. The instrument consists of a light coil that moves through a magnetic field (Figure 51). The electrodynamic principle is governed by the following equation:

$$E = -BLV \times 10^{-8}$$

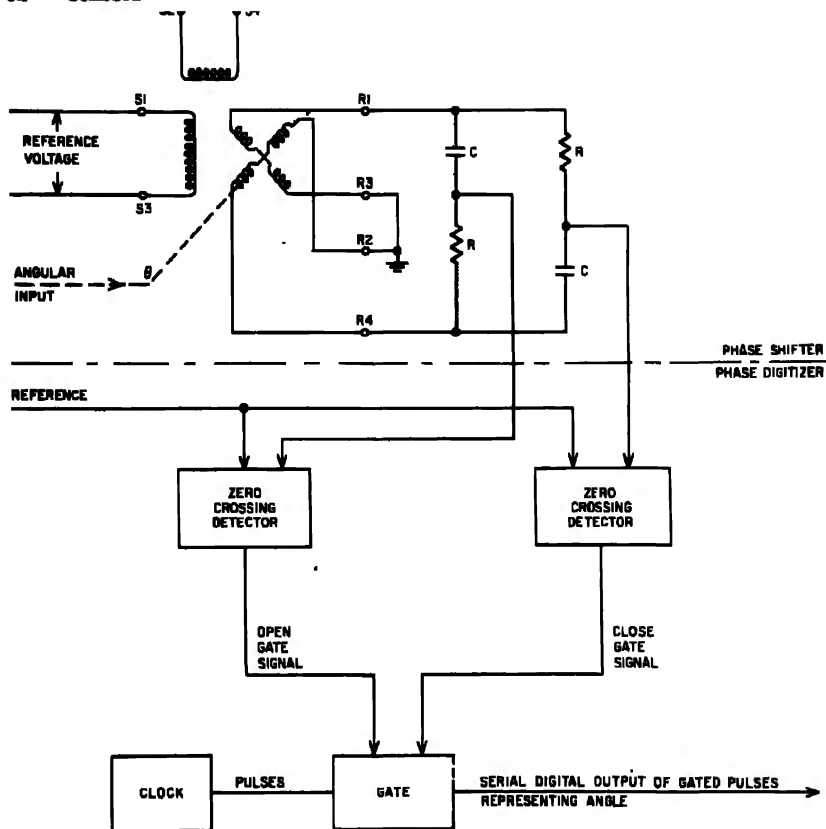


Figure 50. High-accuracy system for angular measurement. (Courtesy of Ford Instrument Division, Sperry Rand Corporation.)

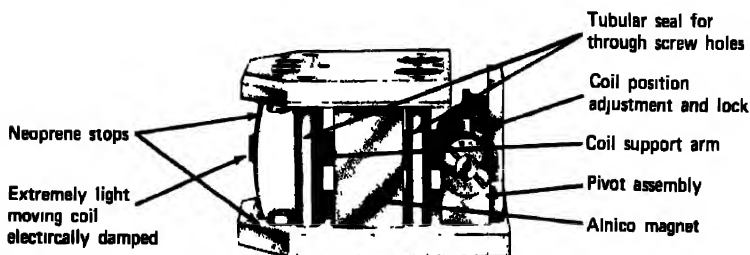


Figure 51. Velocity pickup. (Courtesy of MB Corporation.)

where E = voltage induced in the coil

B = flux density (gauss)

L = total length of the coil (centimeters)

V = relative velocity between the coil and the magnetic field (centimeters per second.)

For a given configuration the output voltage is directly proportional to the velocity of the coil. The vibration table normally moves sinusoidally; consequently the output of the velocity transducer is a sine wave. The motion of the coil is damped using eddy current principles. The support for the moving coil is made of aluminum or some other light, nonferrous material of low resistivity. When it is moved perpendicularly through a magnetic field, a current is generated in the material that is proportional to velocity. These eddy currents set up a magnetic field in a direction opposing the magnetic field that created them. Damping is important because it permits extended flat frequency response and relative immunity to temperature effects. The output voltage may be read on any VTVM or on a velocity meter calibrated directly in inches per second (Figure 52).

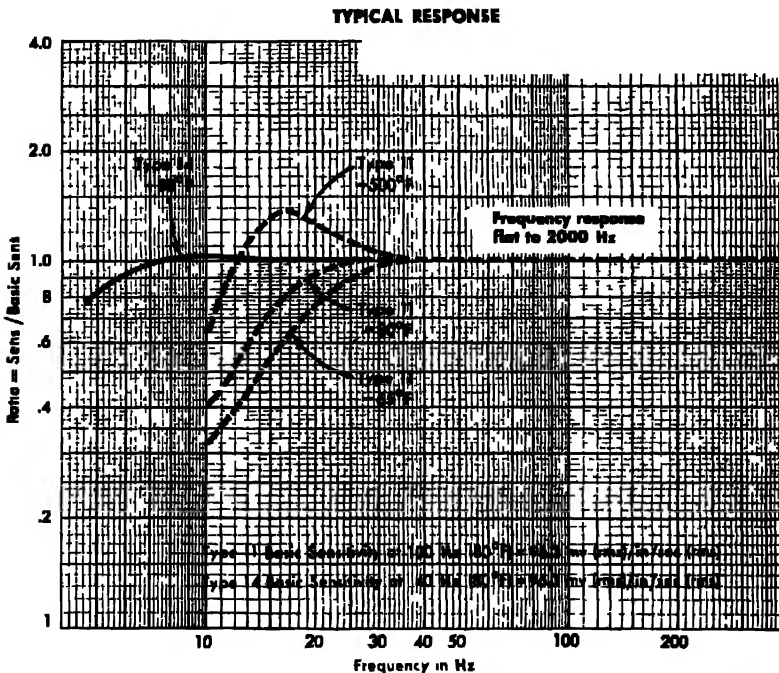


Figure 52. Typical frequency response of vibration pickups. (Courtesy of MB Corporation)

2.9.1. Tachometers

Tachometers are the most widely used angular velocity transducers in systems work. Besides measuring speed accurately, they are used for stabilizing servosystems. Current literature uses the terms tachometer and tachometer-generator synonymously.

The prime function of a tachometer is to produce a voltage proportional to its rotor speed. The chief difference between a tachometer and a squirrel-cage motor is that the former has lower residual or null voltages and less phase shift between the energizing voltage and the output voltage. Other advantages are better linearity, lower inertia and friction, and a more technically refined product. The heart of a tachometer is a drag cup mounted on the rotor shaft that magnetically couples the primary and secondary windings. A drag cup is simply a thin metal cup that rotates in the field of the primary. A current is generated in the cup, which in turn produces a field that couples it to the secondary winding. In this way the secondary is linked to the primary field by way of the drag cup. By Lenz's law a current is generated in the cup which produces its own field that couples the drag cup to the secondary and thus the secondary winding to the primary. When the drag cup is at rest, the tachometer is in essence a transformer with zero coupling. As the speed of the instrument increases from zero, an output voltage is generated that is directly proportional to velocity. In this way the coupling increases between the primary and secondary until maximum rated speed is reached (Figure 53).



Figure 53. Drag-up tachometer. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

There are three broad categories of tachometers:

1. Precision tachometers designed primarily for accurate computational functions. Temperature, frequency, and voltage stability are held to extremely close tolerances by auxiliary compensation controls.
2. Rate generators have high output-to-null ratios and are designed for applications such as rate servos and for providing damping in high-gain servosystems.
3. Damping tachometers have relatively low output-to-null ratios and are designed primarily for damping purposes. They characteristically have very low inertia and low power consumption. Damping tachometers are usually integrally coupled to a low-inertia motor.

In selecting a tachometer the performance features that are of prime importance are accuracy, output, noise, and impedance characteristics. Some of the important criteria in evaluating the performance features are listed below.

Scale factor, S , is defined as follows:

$$S = \frac{\text{in-phase output voltage}}{\text{shaft rpm} \times \text{excitation voltage}}$$

The output of a tachometer is defined as the in-phase open circuit output voltage per unit shaft speed under rated excitation. It is usually rated at 1000 rpm.

Noise in a tachometer at null is the residual voltage existing when the generator is at zero speed. It is of great importance in rate or integrating servos but of lesser importance for damping a position servo. Null voltage is measured with rated excitation and consists of three components: fundamental in-phase voltage, fundamental quadrature voltage, and harmonics.

Fundamental in-phase voltage is the open-circuit output voltage that is in phase with the excitation voltage. It adds voltage to the output when the shaft rotates in one direction and subtracts voltage when the shaft rotation is reversed. It is therefore a serious error at low speeds. The in-phase component is kept to an absolute minimum in tachometers used for rate or integrating servos. Creep of a rate servo with zero input is commonly caused by an excessive in-phase null voltage.

Fundamental quadrature null voltage is in time phase quadrature with the excitation voltage; it lags the excitation frequency by 90° .

The harmonics are higher time-phase components of the fundamental; they are typically the third harmonic.

The amount of null voltage a system can tolerate is different for each application. It is a function of amplifier gain, the ability of the amplifier to

discriminate between quadrature and in-phase signals, and the capability of the amplifier to attenuate higher harmonics.

The impedance characteristics that can cause trouble in a servosystem are variations of magnitude and phase angle when the load and temperature change.

Linearity is the deviation of the in-phase output voltage-shaft speed gradient from a constant value. It is usually expressed as a percentage of the output at 3600 rpm and is the maximum deviation from the best straight line drawn through points from 3600 rpm clockwise to 3600 rpm counter-clockwise. Typical values are 0.03 to 0.1%.

One of the largest errors in tachometers is caused by the effects of temperature changes. Two methods are used to control this problem:

1. Unit temperature control.
2. Temperature compensation.

The circuit shown in Figure 54 is used for unit temperature control. Two thermistors are imbedded in the tachometer windings. The resistors in the bridge are chosen to be equal to the thermistor resistance at the controlled temperature. At any other temperature there will be an error voltage at the input to the magnetic amplifier and the heater will be energized. The normal control temperature is higher than ambient and is typically 125 to 160°F. Because of the size of the amplifier, heater, and power requirements, this method is not always practical.

The temperature compensation method shown in Figure 55 utilizes a

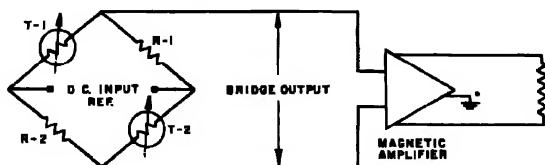


Figure 54. Magnetic amplifier temperature control for tachometers. (Courtesy of Singer-General Precision, Inc. From Reference 16.)

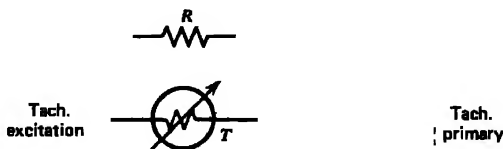


Figure 55. Thermistor compensation for tachometers. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

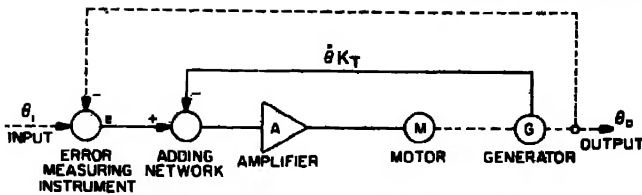


Figure 56. Tachometer used in position servo for damping. The $\dot{\theta}$ = motor speed; K_T = tachometer constant. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

thermistor-resistor network in series with the tachometer primary. Over a range of -15 to 71°C the resistance of the primary circuit may be held within $\pm 0.2\%$. This is normally the preferred method for eliminating temperature errors.

TACHOMETER APPLICATIONS. The most common application for a tachometer is in a position or type 1 servo (Figure 56). In this mechanism the tachometer is used in the feedback loop to provide the electrical equivalent of viscous friction.

Another application is found in velocity or type 2 servomechanism where the tachometer feeds back a voltage proportional to velocity (Figure 57).

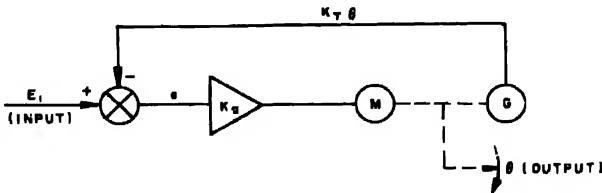
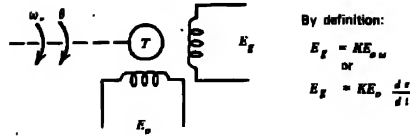


Figure 57. Velocity servo for integration. The figure depicts a simple closed-loop integrator employing a servomotor integrally coupled to a precision AC tachometer. Assuming an amplifier gain K_a approaching infinity, the error signal, e , approaches zero, and the voltage output from the tachometer, $K_T \dot{\theta}$ will approach a value equal to the input signal, E_i : $E_i \rightarrow K_T \dot{\theta}$; $\int E_i dt \rightarrow K_T \int \dot{\theta} dt \rightarrow K_T \theta$. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

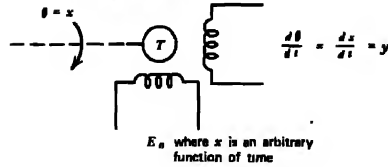
The third application is in integrators. This case is discussed in Chapter 4. For more complete servo analysis see *Servo Mechanism Design* by Chestnut and Mayer (Reference 18).

Other tachometer applications are shown in Figure 58.

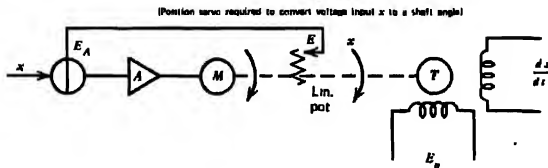
Typical tachometer applications
Fundamental component



Time differentiation
Mechanical input



Time differentiation
Voltage input



System demands $E_A = x$, since $E_A = \theta$

Then $\theta = x$, hence $y = \frac{dx}{dt}$

Differential-variable other than time

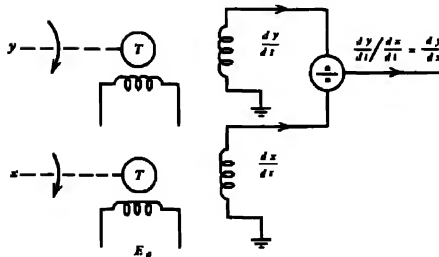


Figure 56. Tachometer applications. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

Hence the output of the 2nd pot. = $\frac{dy}{dt} \times \frac{1}{dx/dt} = \frac{dy}{dx}$

[illegible]
$$\text{or } \theta = \int x \, dt = y$$

99

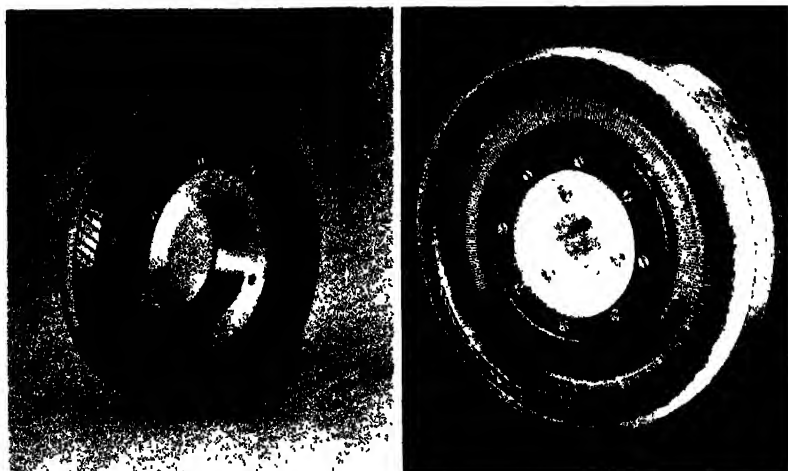


Figure 59. Typical DC tachometer generator rotors. (Courtesy of Inland Motors Division, Kollmorgen Corp., Radford, Va. From Reference 19.)

DC TACHOMETERS. So far only AC tachometers have been discussed. DC tachometers were recently developed to meet requirements in cases where the size and weight of the tachometer are critical. The envelope of DC tachometers is basically different from AC units. They are designed to fit on the customer's shaft; they have no shaft or frame of their own. The unit is generally very thin in relation to its diameter (Figures 59 and 60).

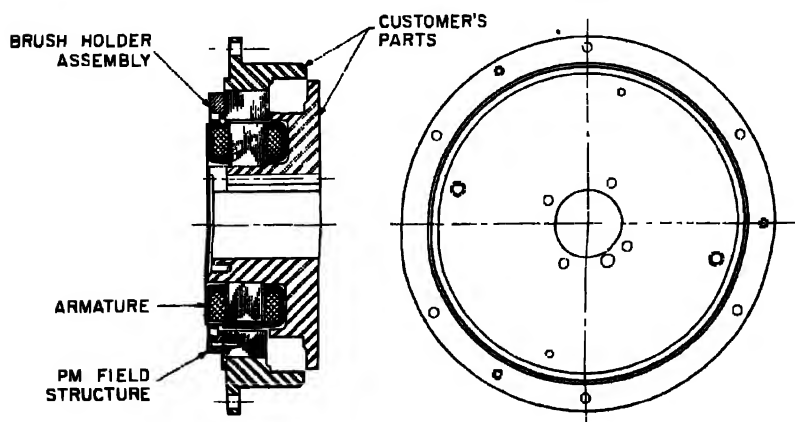


Figure 60. Frameless tach generator with adaptors machined to existing mounting dimensions. (Courtesy of Inland Motors Division, Kollmorgen Corporation. From Reference 19.)

The DC tachometer has a wound armature and a permanent magnet field. When the unit is mechanically driven, it produces a voltage proportional to velocity. This system has the following advantages:

1. High coupling stiffness—since the instrument is directly coupled to the load, there are no gears and resulting backlash.
2. High voltage gradient—the ratio of voltage to rotational speed is high.
3. High linearity and resolution.
4. Fair response.

2.9.1.1. DC Tachometer Applications

Frequently the pancake-style torque motor in standard or modified form is used as a generator to supply a voltage proportional to shaft speed (Figure 61). The transfer characteristics for such a device are derived from

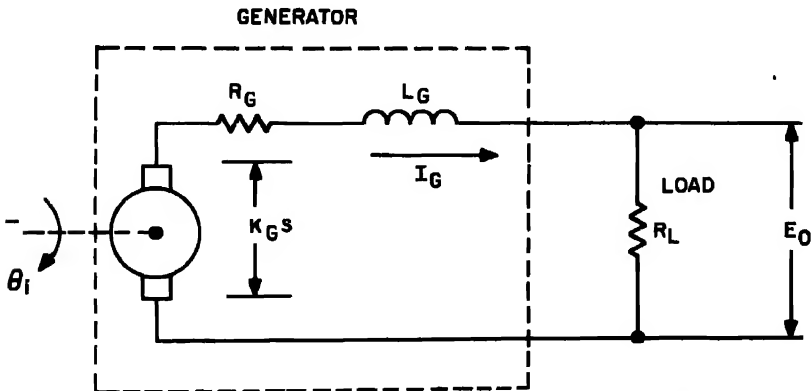


Figure 61. Equivalent circuit of a torque motor used as a tachometer generator, including load. (Courtesy of Inland Motors Division, Kullmorgen Corporation. From Reference 19.)

the equivalent circuit shown. Writing the loop equation, we have $K_G \theta_i s = (R_G + R_L)I_G + L_G I_G s$ and $E_o = I_G R_L$. Combining them to eliminate I_G , we obtain

$$\frac{E_o}{\theta_i} = \frac{R_L}{R_G + R_L} \frac{K_G s}{1 + [L_G / (R_G + R_L)]s} \quad (17)$$

From this transfer function the output is a voltage proportional to the first derivative of shaft position, or velocity, as modified by the electrical lag in the armature circuit.

The effect of rate feedback on the loop dynamics can be investigated

with the block diagram of such a system (Figure 62). From the diagram, the closed-loop transfer function for the inner loop is

$$\frac{\theta_o}{\epsilon_2} = \frac{(K_A/K_B)/[s(1 + \tau_M s)]}{1 + [(K_A/K_B)K_G]/(1 + \tau_M s)}$$

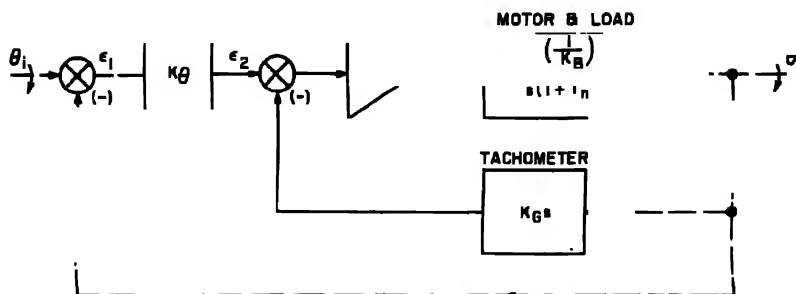


Figure 62. Closed-loop transfer function for a servo with tachometer generator (rate) damping. (Courtesy of Inland Motors Division, Kollmorgen Corporation. From Reference 19.)

Simplifying, we have

$$\frac{\theta_o}{\epsilon_2} = \frac{K_A/K_B}{s(1 + \tau_M s) + (K_A K_G/K_B)s} \quad (18)$$

where K_A = amplifier gain, K_B = motor back EMF, τ_M = motor mechanical time constant, K_G = tachometer constant, ϵ_2 = error signal applied to the amplifier, θ_o = displacement of the output shaft, and s = Laplace notation for the first derivative of angular displacement with respect to time.

2.9.2. Electro-Magnetic Velocity Pickup

Another interesting and inexpensive velocity transducer is the magnetic velocity pickup. The pickup head consists of a magnet and coil encapsulated in a stainless steel shell. The pickup generates a voltage output when any ferromagnetic material enters the magnetic field, since the flux through the coil is increased. The amplitude of the voltage is proportional to the speed at which the magnetic material moves through the field. Practical setups include a gear fastened to a rotating shaft and a pickup positioned about 0.020 to 0.070 in. from the gear. The output of the device may be monitored on a scope, recorder, or a specially calibrated frequency meter. The technique is largely independent of the wave shape generated (Figure 63).

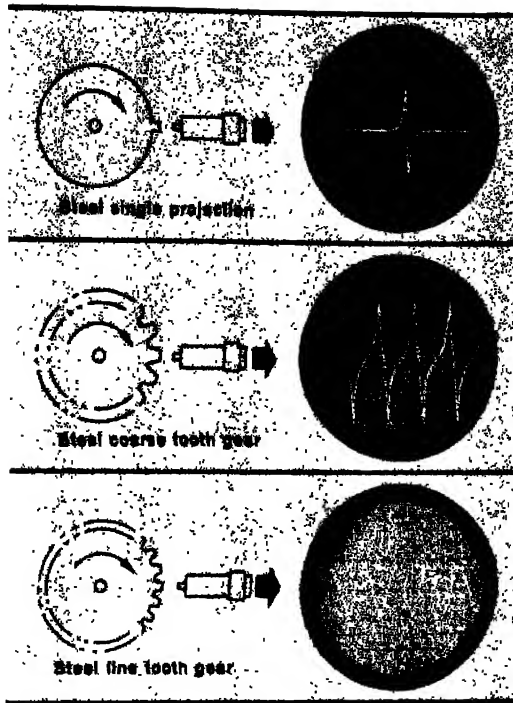


Figure 63. Output waveforms from a magnetic velocity pickup. (Courtesy of Electro-Products Corporation.)

2.10. ACCELERATION TRANSDUCERS

All acceleration transducers, or accelerometers, are governed by Newton's second law, which states that force is equal to mass times acceleration.

where F = force (dynes)

M = mass (grams)

A = acceleration (centimeters per second)

In the conventional English system:

$$F = \frac{W}{G} A$$

where F = force (pounds)

W = weight (pounds)

G = acceleration of gravity [32.2 ft/(sec)/(sec)]

A = acceleration of the mass [ft/(sec)/(sec)]

Although many design problems involve linear acceleration, a greater number entail acceleration and deceleration as the result of periodic motion. A practical simulation of most repetitive motion is the sine wave. The relation between displacement, velocity, and acceleration is as follows:

$$\text{displacement:} \quad x = A \sin \omega t \quad (19)$$

$$\text{velocity:} \quad v = \omega A \cos \omega t \quad (20)$$

$$\text{acceleration:} \quad a = -\omega^2 A \sin \omega t \quad (21)$$

where A = maximum displacement amplitude

ω = frequency of the forcing function (radians per second)

t = time (seconds)

Few motions in nature actually approximate pure sine waves. There are usually at least a few harmonics present. Assume that one such waveform includes a fundamental and third harmonic. Assume also that the peak amplitude of the third harmonic, Q , is one-third the amplitude of the fundamental P (Figure 64).

$$x = P \sin \omega t + Q \sin 3\omega t \quad (22)$$

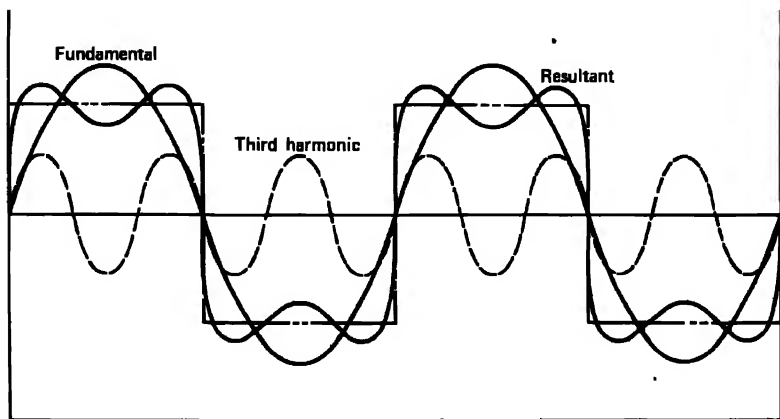


Figure 64. Steady-state-complex periodic motion (fundamental and third harmonic of a square wave). (Courtesy of Columbia Research Laboratories, Inc. From Reference 20.)

For a greater number of harmonics P , the following generalization is true:

$$x = P_0 + \sum_{n=1}^{n=P} (P_n \cos n\omega t + Q_n \sin n\omega t) \quad (23)$$

Random vibration analysis utilizes a statistical approach to problems where acceleration varies with time in a nonperiodic manner. An example of random vibration is shown in Figure 65, where the magnitude of the peaks

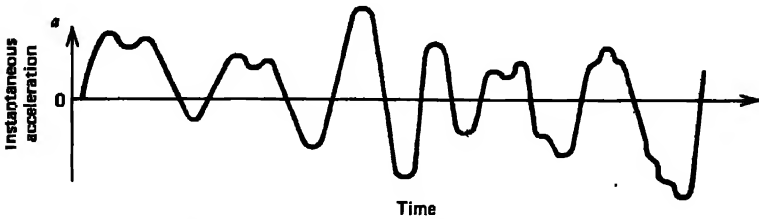


Figure 65. Acceleration-time curve of typical random motion. (Courtesy of Columbia Research Laboratories, Inc. From Reference 20.)

and the period of time between waves vary irregularly. Motion of this type is the result of a very large number of events occurring by chance. The waveform can be considered as composed of an infinite number of closely packed sinusoidal curves, each cycling at its own frequency with different amplitudes that vary with time. (See Reference 20.)

When the probability of the occurrence of random events must be predicted, probability theory is used. From the nature of the continuous frequency spectrum of random vibration it is reasonable to assume the instantaneous accelerations in a given frequency band have a normal probability distribution about zero as a mean. A plot of the relative frequency of occurrence, or probability density, of an acceleration which follows the normal probability law is shown in Figure 66. The abscissa a/\bar{a} is the ratio of the instantaneous acceleration, a , to the standard RMS acceleration \bar{a} . The area under the curve between any two limits is the probability that a typical of a/\bar{a} will occur within these limits. For the case $a/\bar{a} = 1$, the instantaneous acceleration, a , is equal to or less than the standard deviation 68.3% of the time.

The mean acceleration density, $G(f)$, the spectral distribution as a function of time, is defined as follows:

$$G(f) = \frac{\lim_{B \rightarrow 0} [\bar{a}]^2}{B} \quad (24)$$

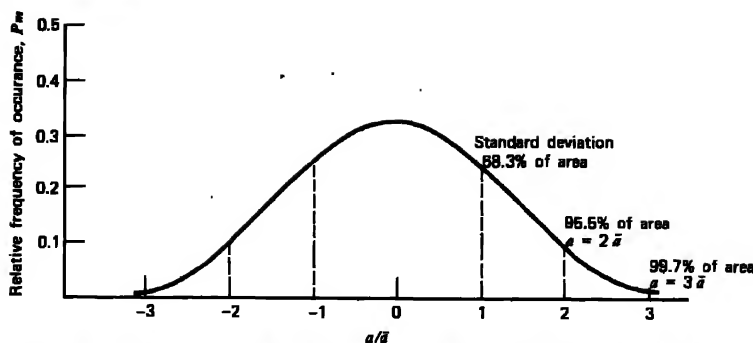


Figure 66. Gaussian or normal distribution curve is useful inasmuch as the area under the curve P_m between any two limits is the probability that a typical value of a/\bar{a} will fall within these limits. Thus in the case of $a/\bar{a} = \pm 1$ we can say the instantaneous acceleration a is equal to or less than the standard deviation 68.3% of the time. (Courtesy of Columbia Research Laboratories, Inc. From Reference 20.)

where \bar{a} = RMS of random accelerations

B = bandwidth (hertz)

$G(f)$ = mean acceleration density (g^2/hertz)

g = acceleration due to gravity

Acceleration density is sometimes referred to as the mean square acceleration, spectral density, mean square acceleration per cycle, or the power density.

Since acceleration density is a function of frequency, the RMS acceleration in the band between f_1 and f_2 can be calculated from the following:

$$[\bar{a}]^2 = \int_{f_1}^{f_2} G(f) df \quad (25)$$

A random vibration having a constant acceleration density with frequency is referred to as white noise. The acceleration density is independent of bandwidth and the equation simplifies to

$$G_0 = \frac{[\bar{a}]^2}{B} \quad (26)$$

where G_0 = white noise acceleration density.

$$\bar{a} = \sqrt{BG_0} \quad (27)$$

Example. Find \bar{a} , when $B = 2000$ Hz and $G_0 = 0.5g^2/\text{Hz}$.

$$\bar{a} = \sqrt{2000 \times 0.5} = 31.7g \text{ (RMS)}$$

2.10.1. Selection of Accelerometers

Selection of the proper acceleration transducer is largely determined by the bandwidth of the function being monitored. Accelerometers' performance may be mathematically approximated by a second-order equation.

$$P_0 \sin \omega t = M\ddot{X} + F\dot{X} + KX \quad (28)$$

where M = mass of the seismic element
 F = damping coefficient
 K = spring constant
 X = displacement of the seismic element

$P_0 \sin \omega t$ = external forcing function of peak amplitude P_0

The undamped natural frequency is

$$\omega_n = \sqrt{K/M} \quad (29)$$

The damped natural frequency is

$$Q = \sqrt{K/M - [(F/2M)]^2} \quad (30)$$

For accurate work the natural frequency of the accelerometer should be at least five times the highest frequency being measured (Figure 67). The relative response of an accelerometer, R , is defined as follows:

$$R = \frac{1}{1 - (f/f_n)^2} \quad (31)$$

where f = forcing frequency (hertz)

f_n = natural frequency of the accelerometer (hertz)

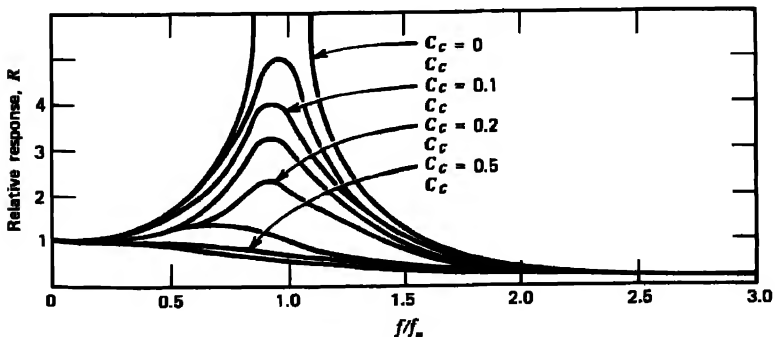


Figure 67. Relative response versus ratio of forcing frequency to natural frequency. c/c_c = damping ratio. (Courtesy of Columbia Research Laboratories, Inc. From Reference 20.)

Example. For a 4% error find the ratio of f/f_n required.

$$1.04 = \frac{1}{1 - (f/f_n)^2}$$

$$f/f_n = 0.2$$

The effect of damping is assumed to be negligible.

Accelerometers fall into two general categories: navigational instruments and laboratory instruments. Navigational units include the following:

1. Vertical references.
2. Force balance accelerometers.
3. Integrating accelerometers.
4. Vibrating string accelerometers.

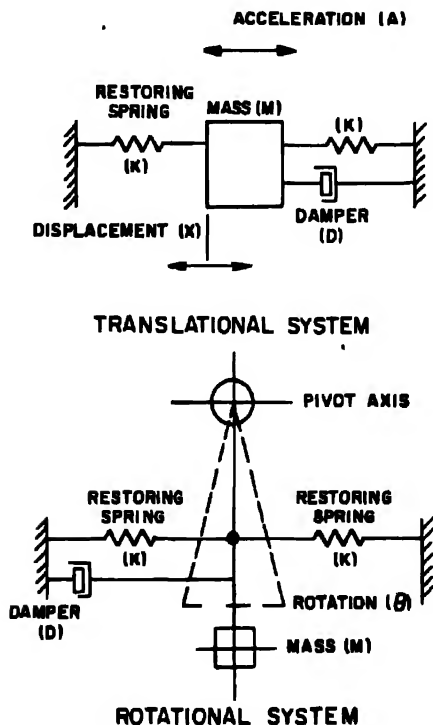


Figure 68. Vertical reference accelerometers. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

Laboratory units include:

1. Piezoelectric accelerometers.
2. Piezoresistive accelerometers.

VERTICAL REFERENCES. Vertical reference accelerometers are used in inertial guidance platforms for determining the direction of a "plumb line" to the center of earth. A vertical reference accelerometer is an instrument consisting of a damped spring-mass, second-order, single-degree of freedom system (Figure 68). The transfer function is shown in Figure 69. Vertical

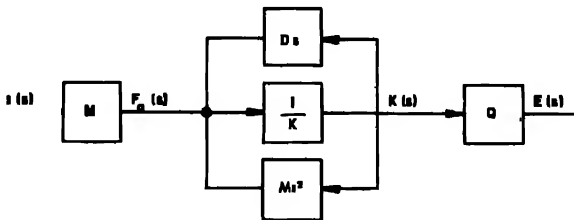


Figure 69. Accelerometer transfer function. a = Input acceleration, M = mass, F_a = acceleration force, D = damping (viscous), K = spring constant, Q = transducer scale factor, s = Laplace operator, and E = output voltage. Overall transfer function = $E/a(s) = Q/(s^2 + Ds/M + K/M)$. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

reference accelerometers are sometimes classified according to the transducer used to convert mass motion to an output voltage. An LVDT or synchro is often used for AC systems and a potentiometer for DC systems. Other units employ piezoelectric devices, strain gages, and vibrating wires.

A vertical reference device acting as a "plumb line" on the rotating nonspherical earth assumes the direction in which gravity acts on the rotating earth. Since gravity is a resultant between the earth's mass attraction force and its rotational centripetal force, and because the earth is not a perfect sphere, geometric vertical does not coincide with gravity vertical (Figure 70).

Vertical sensors are used as an auxiliary device to turn on and off critical gravity-sensitive units. The instrument is generally considered to be a special type of accelerometer. The construction is basically a mass at the end of a rod or wire, damped with a fluid or magnetically restrained. The rod may be suspended on jewel bearings or as a torsion bar. Normally, no spring is attached to the mass, so that a relatively small displacement from vertical will cause the mass to operate a switch of some sort. One type of

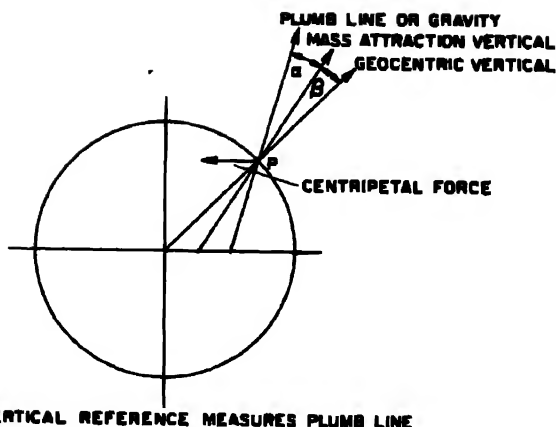


Figure 70. Vertical reference. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

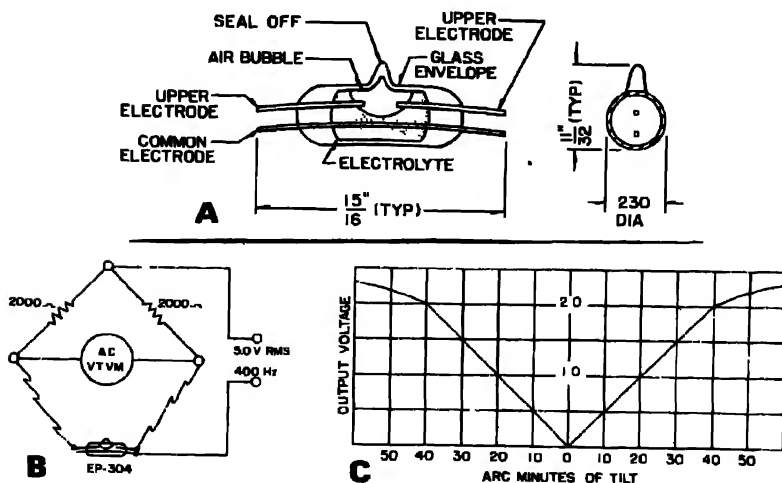


Figure 71. Bubble-type gravity sensor. (a) Cutaway view showing construction features. (b) Resistive bridge circuit used to test and determine operating characteristics. (c) Typical output sensitivity when operated in the standard test circuit at room ambient environment. (Courtesy of Hamlin, Inc.)

seismic system is simply a bubble of conductive material, such as mercury, that shorts two terminals in the vertical position (Figure 71). The output is interrupted when the bubble is displaced. Other units use a synchro-type pickoff that produces a linear variation in output up to a certain number of degrees. The normal range of these devices is from a few minutes of arc to about 10° . Repeatability is usually excellent.

FORCE BALANCE ACCELEROMETERS. A force balance accelerometer is similar to the vertical reference accelerometer with the important exception that the motion of the pendulum, or mass, is restricted to a very small angle—typically under 0.1° . This is to prevent cross-coupling or sensitivity to accelerations along an axis at right angles to the primary axis. Since cross-axis sensitivity is proportional to the sine of the deflection angle of the pendulum, the motion of the seismic element must be held to small values for an accurate system. The force balance accelerometer uses an electrical “spring” to hold the displacement to very small magnitudes (Figure 72).

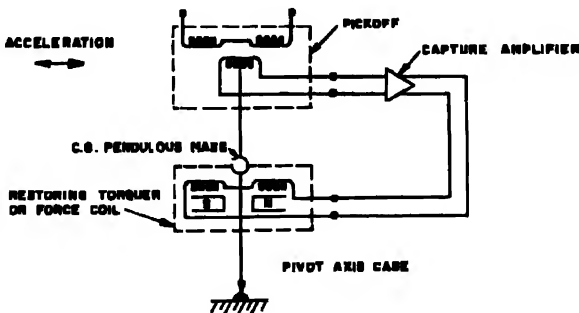


Figure 72. Force-balance accelerometer. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

The instrument is equipped with a pickoff device that detects motion of the mass. Pickoff output resulting from mass motion is directed to a high gain servoamplifier that energizes a force balance coil or torquer, that produces sufficient force to return the mass to its null position. The output of the amplifier is a voltage proportional to the acceleration being measured. The transfer function is shown in Figure 73.

PENDULOUS INTEGRATING ACCELEROMETERS. This instrument is a single-axis gyro with an unbalanced mass, m , positioned at a distance, l , from the center of the gyro wheel. This unbalance, or pendulosity, is the product of

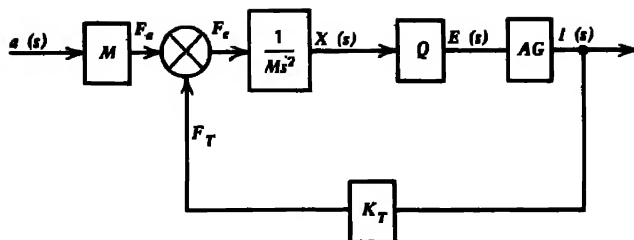


Figure 73. Transfer function of force-balance accelerometer. a = acceleration, M = pendulous mass, F_a = force due to acceleration, F_T = restoring force, F_e = error, x = linear displacement of mass, A = amplifier gain, Q = pickoff scale factor, G = transfer function of shaping or stabilization networks and is in the form of a typical lag-lead network such as $G = (T_1s + 1)(T_2s + 1)/(T_3s + 1)(T_4s + 1)$ where T_1 , T_2 , T_3 , and T_4 are time constants and s is the Laplace operator. The overall transfer function $= 1/a(s) = QAG/(s^2 + QAGK_T/M)$. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

the unbalanced mass multiplied by its moment arm from the gyro's precession axis (Figure 74):

$$\text{pendulosity} = ml \quad (32)$$

An acceleration along the input axis causes motion about the precession axis because of the mass unbalance. The gyro is mounted in a gimbal, which in turn is supported by precision bearings, mounted in the case of the device. The gyro is part of a null-seeking servosystem that measures the displacement of the gyro about its input axis by means of a pickoff; these data are directed to a high gain amplifier that energizes a torquer to rotate the gimbal in a direction that brings the system back to a null position. The angular motion of the gimbal is equivalent to the integral of acceleration, or velocity.

The torque due to acceleration is

$$T_a = mal$$

where a = acceleration

l = distance from the mass to the center of the gyro wheel

The restoring torque is

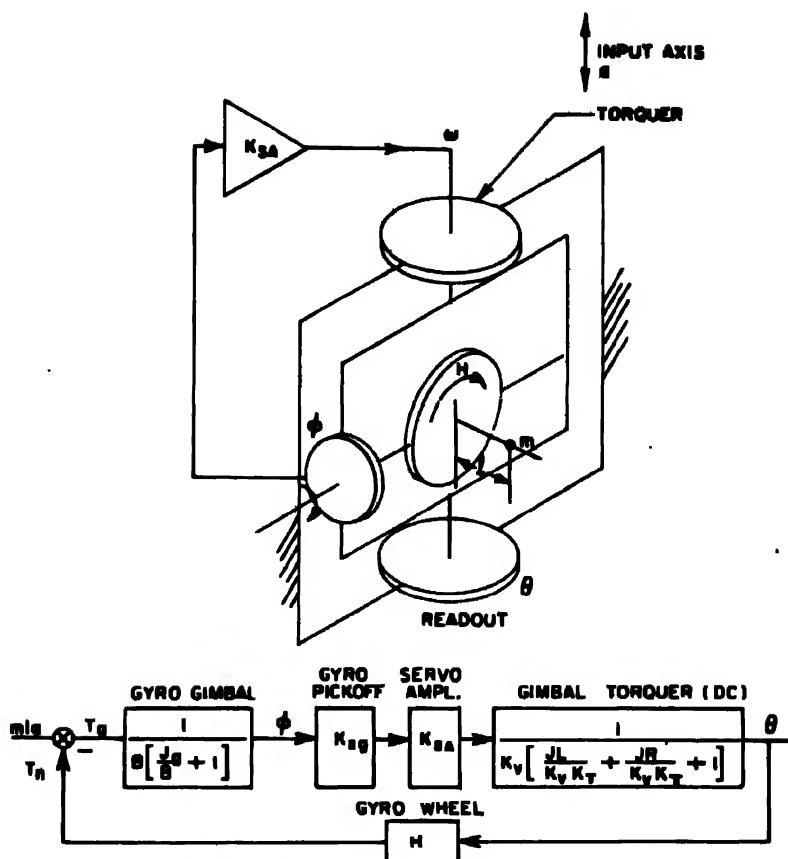
$$T_r = \omega H$$

where ω = input rate to the gyro or the gimbal angular velocity

H = gyro angular momentum

For null conditions, $T_r = T_a$

$$\omega = \left(\frac{ml}{H} \right) a$$



B = GYRO PRECESSION AXIS DAMPING dyne cm sec
 J_g = MOMENT OF INERTIA ABOUT PRECESSION AXIS gm cm²
 L = MOTOR ARMATURE INDUCTANCE henry
 R = MOTOR ARMATURE RESISTANCE ohms
 H = GYRO ROTOR ANGULAR MOMENTUM gm cm²/sec
 J_m = MOMENT OF INERTIA OF ACCELEROMETER ABOUT INPUT AXIS gm cm²
 K_v = MOTOR BACK emf volt/rad/sec
 K_T = MOTOR TORQUE CONSTANT dyne cm/amp

Figure 74. Pendulous integrating gyroscope accelerometer. (Courtesy of Singer-General Precision, Inc. From Reference 14.)

When both sides of the equation are integrated with respect to time, we have

$$\theta = \left(\frac{mI}{H} \right) v \quad (33)$$

This equation states that the output angle of the gimbal, θ , is proportional to the velocity of the accelerometer, V , in space.

The prime applications for integrating accelerometers are in inertial guidance platforms.

A much simpler type of pendulous accelerometer is the unit shown in Figure 75. The instrument consists of a cylinder, mounted on jewel-type

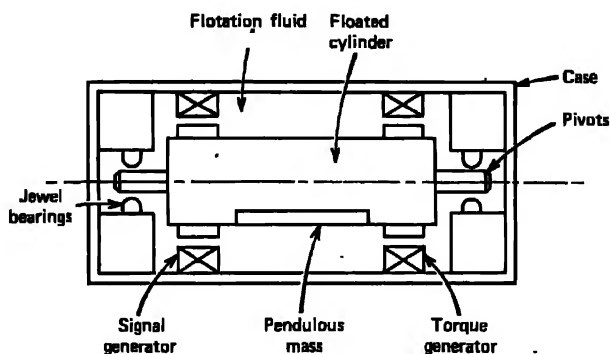


Figure 75. Pendulous floated accelerometer.

bearings, with an intentional mass unbalance built into it. Along the axis of rotation is a pickoff and a torquer. The cylinder of the float is immersed in flotation fluid. Rotation of the float is held under 1° by the servosystem. These units are capable of measuring $1 \times 10^{-4} G$ s and are virtually immune to shock.

VIBRATING BEAM ACCELEROMETERS. Vibrating beam or vibrating string accelerometers are capable of measuring accelerations as low as $1 \mu g$. They consist of a pair of quartz beams, mounted in tension and connected to pivoted seismic masses. Since the natural frequency of the beam is a function of its tension, an input acceleration applied along the longitudinal axis of the device causes an increase in tension in one beam and a decrease in the other. Therefore each beam vibrates at a different frequency (Figure 76). It is important to note that each beam is loaded by a separate inertial mass. The beams are excited to vibrate at their natural frequency by an integral electronic circuit. The difference in frequencies of the two beams is used to

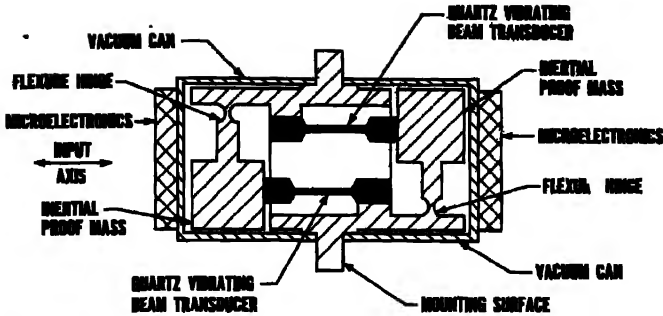


Figure 76. Vibrating beam accelerometer. (Courtesy of Singer-Precision General, Inc. From Reference 14.)

modulate an output voltage that is the analog of acceleration. The signal is then demodulated or detected by appropriate circuitry.

The frequency of oscillation for each beam may be expressed as follows:

$$f_1^2 = \frac{T_1}{4\delta L^2}, \quad f_2^2 = \frac{T_2}{4\delta L^2} \quad (34)$$

where f_1 = frequency of beam 1
 f_2 = frequency of beam 2
 T_1, T_2 = tension in the respective beam
 δ = mass per unit length of the beam
 L = length of the beam

The difference in tension is proportional to mass times acceleration.

$$\begin{aligned} f_1^2 - f_2^2 &= \frac{T_1 - T_2}{4\delta L^2} = \frac{ma}{4\delta L^2} \\ [f_1 - f_2][f_1 + f_2] &= \frac{ma}{4\delta L^2} \\ f_1 - f_2 &= \left[\frac{m}{(f_1 + f_2)4\delta L^2} \right] a \end{aligned} \quad (35)$$

Acceleration is proportional to the difference of the beam frequencies.

These units lend themselves very well to digital techniques, since they measure acceleration by counting the difference in vibration frequency. As in all crystal units, stability is excellent.

PIEZOELECTRIC ACCELEROMETERS. Piezoelectric accelerometers are unique because they generate their own voltage without any auxiliary power

supply. The heart of the instrument is a disk that produces voltage when subjected to pressure. The seismic element is simply a mass that applies load to the piezoelectric crystal when the accelerometer is exposed to acceleration. The inherent simplicity and ruggedness of this instrument has made it standard equipment in most testing labs. Figure 77 shows the

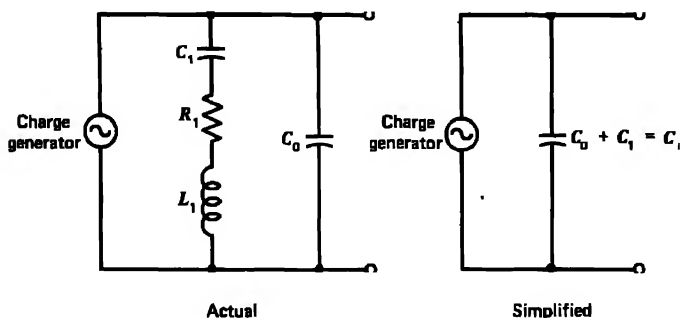


Figure 77. Equivalent circuit piezoelectric transducer. (Courtesy of Endevco Corporation. From Reference 22.)

equivalent electrical circuit. It is essentially a charge generator in parallel with a capacitor. The output voltage, E , is equal to the charge generated divided by the transducer capacitor or $E = Q/C_p$.

Most manufacturers supply these accelerometers calibrated with 100 pF in parallel with the output. This represents the capacitance of the connecting cable and the input capacitance of the associated amplifier. If the cable and amplifier capacitance is substantially different from 100 pF, corrections must be made. The following equation is applicable:

$$E = E_c \frac{(C_p + 100)}{C_p + C_t} \quad (36)$$

where E = sensitivity of the accelerometer for any capacitive loading (millivolts RMS per single amplitude G , or millivolts per peak G)

E_c = factory-supplied calibration

C_p = capacitance of the accelerometer (picofarads)

C_t = externally supplied capacitance (cable and amplifier)

Example. Assume that $E_c = 20$, $C_p = 600$, and $C_i = 200$. Then

$$E = 20 \left[\frac{600 + 100}{600 + 200} \right] = 17.5 \text{ mV RMS/peak } G$$

It is important to note that piezoelectric accelerometers have no response to steady-state acceleration. Their low frequency response is a function of the RC time constant of the accelerometer and the input impedance of the matching elements (Figure 78).

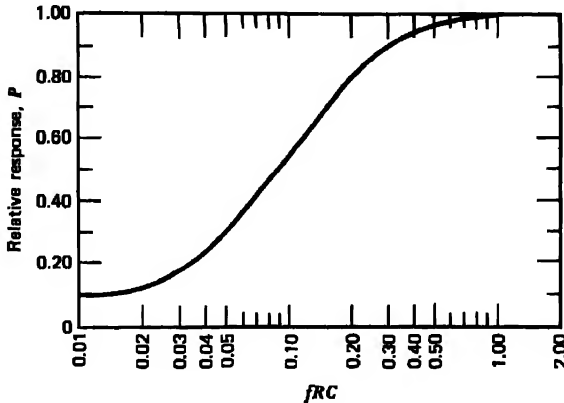


Figure 78. Low-frequency response versus loading. f = Frequency response (hertz), R = input impedance of the amplifier (ohms), and C = capacitance of accelerometer and cable (farads). (Courtesy of Endevco Corporation. From Reference 22.)

For example, assume that it is desirable to have a low-frequency response of 10 Hz. The input impedance of the amplifier (FET input) is 100 M Ω . The total capacitance of the accelerometer, cable, and amplifier is 600 μ F (Figure 78).

$$fRC = 10 \times 100 \times 10^6 \times 600 \times 10^{-12} = 0.6$$

Using Figure 78, the relative response is 0.98. This means that the response will be down 2% at 10 Hz.

The high-frequency response of a piezoelectric accelerometer may be expressed as follows:

$$P = \frac{1}{1 - \beta^2} \quad (37)$$

where P = relative response

β = ratio of the forcing frequency to the natural frequency

A normal rule of thumb is to select an accelerometer with a natural frequency at least five times the highest forcing function anticipated; then $B = \frac{1}{5}$ and $P = 1.04$ or 4% higher than the ideal.

The temperature sensitivity of standard units varies from $\pm 10\%$ over a range of -65 to $+230^\circ\text{F}$, to $\pm 1\%$ for special crystals (Figure 79a).

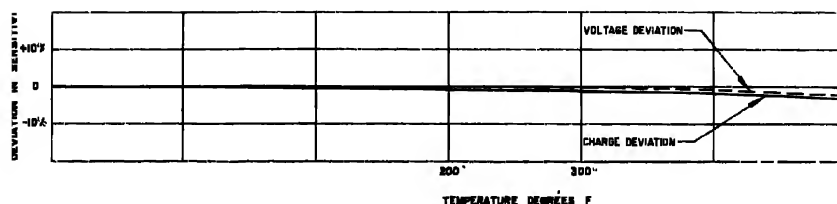


Figure 79a. Deviation of transducer sensitivity with temperature. (Courtesy of Endevco Corporation. From Reference 22.)

Manufacturers have different names for each generic type of crystal; consequently this point must be checked before purchasing the instrument. Typical sensitivity of a piezoelectric unit is 0.1 to 50.0 mV/G.

PIEZORESISTIVE AND STRAIN GAGE ACCELEROMETERS. These instruments are used when DC to moderately high frequencies are to be measured—typically 0 to 2000 Hz. The comments made in Section 2.1.2 are equally applicable here. The chief difference is that the units are calibrated in terms of acceleration instead of displacement. They are often calibrated on a centrifuge to obtain a plot of output voltage versus acceleration.

SHOCK TESTING. Shock testing imposes special problems that are not encountered in normal vibration work. Some of the problems are:

1. High g levels.
2. Very wide frequency response.
3. Transient response characteristics of instrumentation.
4. Repeatability of test conditions.

A shock pulse is usually shaped in the form of a half sinusoid, sawtooth, triangular or square wave. In some cases the shock pulse is specifically shaped to contain a desired amount of energy in certain bands. Most shock pulses contain essentially all frequencies from DC to well above 10 kHz. For shock work the entire system must have a low frequency response 50 times greater than the period of the fundamental frequency for accuracy of 2%.

Figure 79b shows the results of two instrumentation systems, one with frequency response to 2 Hz and the other with response 3 dB down at 25 Hz. Note that the instrumentation system without flat low-frequency response completely distorts the acceleration time curve.

Pulse Width of 100 g Half-Sine Wave Pulse (No negative acceleration)	Indicated Accelerometer Peak Value Frequency Response		Indicated Negative Accelerometer Undershoot Frequency Response	
	3 dB down at 25 Hz	2 Hz to 10 kHz	3 dB down at 25 Hz	2 Hz to 10 kHz
20 msec	0 g	96 g	Unknown	-6 g
11 msec	10 g	98 g	Unknown	-4 g
6 msec	51 g	99 g	-52 g	-2 g
2 msec	75 g	100 g	-40 g	0
1 msec	87 g	100 g	-26 g	0
500 μ sec	91 g	100 g	-18 g	0

Figure 79b. Piezoelectric accelerometer performance parameters. (Courtesy of Endevco Corporation, Pasadena, California. From Reference 22.)

2.11. PRESSURE TRANSDUCERS

Pressure transducers are fundamentally displacement transducers connected to a force-summing device. A force-summing device converts total force to a precise deflection by integrating pressure over a given area. The deflection is converted to a voltage by a displacement transducer. All the generic characteristics of displacement transducers will be found in pressure transducers.

The three principal force-summing devices are:

1. Bourdon tubes.
2. Bellows.
3. Diaphragms.

A Bourdon tube is a C-shaped tube that tends to flatten out when it is exposed to pressure. The application of internal pressure tends to cause the oval cross section to change to a round section and to unroll or straighten the tube. This motion is transmitted by linkages to a potentiometer, LVDT, or to other displacement transducers (Figure 80). Bourdon tubes are relatively inexpensive instruments that provide 1 to 5% accuracy. The position transducer in this instrument is usually a potentiometer as it combines low cost with reasonable accuracy. The more expensive units

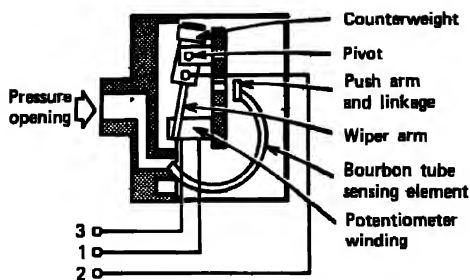


Figure 80. Bourdon tube pressure transducer. (Courtesy of Benwill Publishing Company. From Reference 1.)

usually use a LVDT to sense the motion of the tube. The chief problem associated with Bourdon instruments is hysteresis, or nonrepeatable readings when cycled with ascending and descending pressures. This problem is largely a function of the material used for the tube and how it is heat-treated. The best results have been obtained with beryllium copper. Recently, fused quartz Bourdon capsules have become available. Other normal problems include friction and wear in the connecting linkages. This class of device is not designed for fast frequency response or small envelopes. The basic use is in laboratory and plant control systems. The useful range is from -100 to 500°F with pressures up to 10,000 psi.

Bellows pressure transducers provide a method of obtaining a relatively large displacement for a given pressure variation. Pressure applied to the face of the bellows causes it to contract and move an attached displacement indicator. The advantage over Bourdon types is lower hysteresis and smaller overall size. When the bellows is fabricated from NI-SPAN-C alloy, a zero thermoelastic coefficient is possible (zero change in spring constant with temperature) (Figure 81). This instrument can be used as a differential gage by venting one side of the bellows to a reference pressure. Position transducers used in these devices are linear potentiometers or LVDTs. The former is cheaper and has high output voltage; the latter has greater resolution, higher cost but lower output voltage, and requires AC excitation. A major disadvantage of the linear potentiometer is the wear and noise associated with sliding contacts. Many precise servosystems, such as aircraft instruments, use a buzzer or vibrator to provide a slight amount of vibration that helps minimize static friction. Although this technique helps to reduce friction between pivots and jewel bearings, it has the undesirable effect of causing a potentiometer arm to oscillate back and forth over a very small region, greatly accelerating wear. The LVDT does not suffer from this malady and also has an essentially constant output impedance regardless of core position.

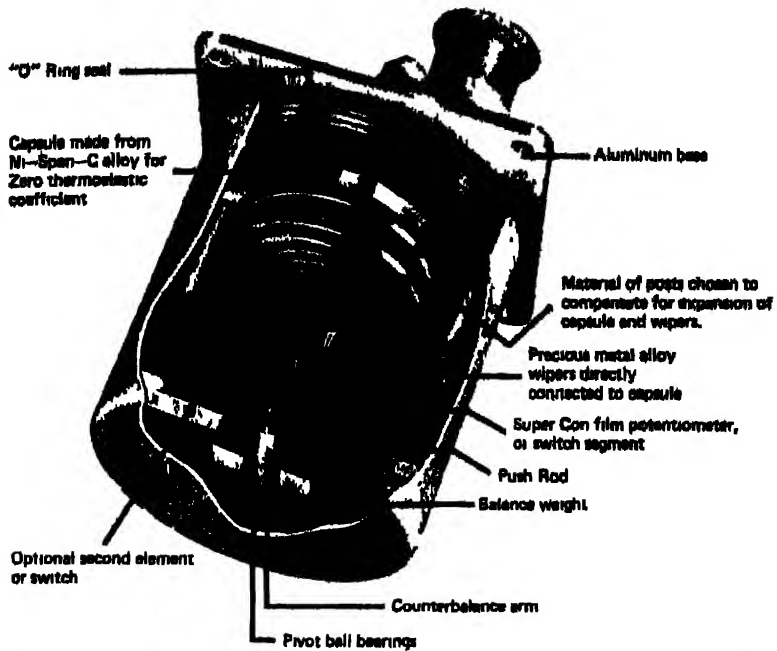


Figure 81. Bellows-type pressure transducer. Multiple-finger wipers directly fastened to the pressure-sensing element ride on the mirrorlike surface of a film-type electrical resistance element, thus converting the motion of the pressure-sensing element into voltage output (Courtesy of Computer Instrument Corporation)

A direct connection between the bellows and the moving part of the position transducer eliminates linkages so that errors due to friction and wear are minimized. The accuracy of bellows pressure transducers are typically $\frac{1}{2}$ to 2%.

Diaphragm-type pressure transducers are the most accurate, have the highest capacity, and are the most versatile commercial units available. The diaphragm, on exposure to pressure, deflects, and the motion is measured by some form of displacement device. The most common ones are strain gage bridges, LVDTs, and capacitive pickups. An interesting example of a diaphragm device is the differential diaphragm pressure gage shown in Figure 82. The diaphragm is made of a magnetic material and mounted between two pickoffs. The pickoffs monitor the gap between their pole pieces and the diaphragm. When the gap changes due to pressure applied to the diaphragm, the output voltage of the device varies. Since one pickoff

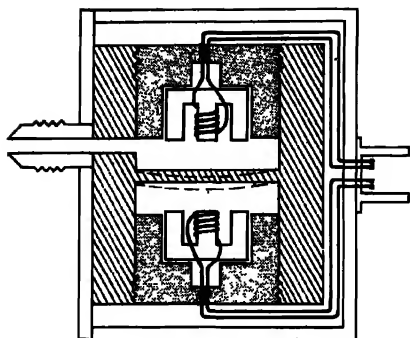


Figure 82. Diaphragm-type pressure transducer. This is a differential pressure sensor whose diaphragm deflection is proportional to the difference in pressure in two cavities. The diaphragm is mounted between two LVDTs that respond to the gap between the face of the pressure-sensing diaphragm and the pole pieces of the magnetic pickoff. Pressure applied to the diaphragm changes the gap, thus changing the output voltage of the device. This instrument, powered by DC, also produces a variable DC output, requiring auxiliary circuitry. (Courtesy of Bourns, Inc.)

is mounted on the side of the diaphragm vented to a reference pressure, and the other pickoff monitors the unknown pressure, the difference in pickoff voltage represents the required differential pressure. This arrangement makes the instrument self-compensating for temperature variations. The circuitry associated with this transducer is particularly interesting. The unit is energized by DC that powers an oscillator at about 10 kHz. The oscillator excites the LVDT, and its output is demodulated and amplified. The output is typically ± 5 to ± 10 V DC. The DC to AC to DC technique has the following advantages:

1. A relatively small battery supply can be used to drive an AC instrument.
2. The sensitivity of the LVDT makes possible the use of a relatively stiff diaphragm with a high natural frequency and good frequency response. The response to shock and vibration is low.
3. The design is suitable for solid-state circuitry and can be made in a small package.

Another interesting variation on diaphragm-LVDT pressure transducers is the Applanation transducer (Figure 83). This device is used to monitor physiological pressure and other medical data. It contains a plunger that is brought into contact with the patient's skin. It is depressed by a fixed amount and held steady by hand pressure. The plunger then moves a minute amount because of pressure changes caused by blood pressure and

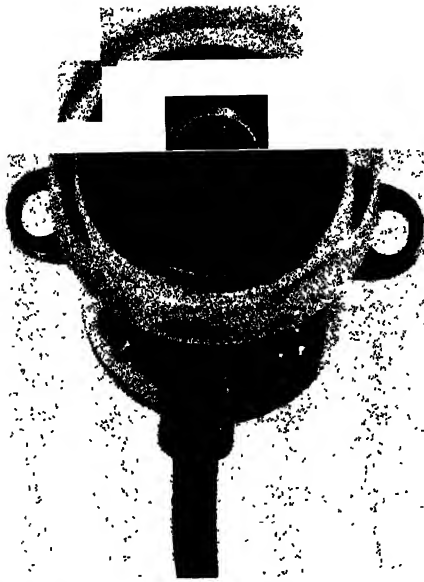


Figure 83. Applanation transducer. (Courtesy of Hewlett-Packard Corporation.)

other physiological factors. The motion of the plunger is monitored by an LVDT using amplitude modulation techniques (DC to DC circuit described above). The instrument has a natural frequency of 300 kHz.

The vast majority of diaphragm pressure transducers use strain gages to monitor the deflection of the diaphragm. This is because they are inexpensive, small, and are reasonably accurate. The strain gages are bonded to the diaphragm and, when pressure is applied, the diaphragm deflects, subjecting the gage to some deformation. All other characteristics are essentially the same as those discussed in Section 2.1.2.

Perhaps the most highly refined diaphragm pressure transducer is the modern altimeter that has been continually improved for the past forty years (Figure 84). Paradoxically, the motion of the diaphragm, which is proportional to altitude, is monitored by a completely mechanical system: The diaphragm is connected to an elaborate gear train and temperature-compensating system that moves a dial. The reason for a completely mechanical approach is that the instrument must function through any type of power failure.

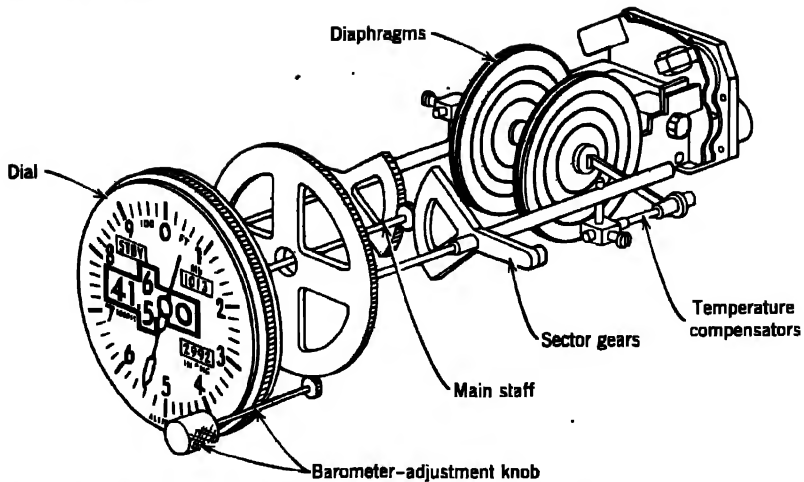


Figure 84. Altimeter. (Courtesy of Kollman Instrument Company.)

Pressure switches employ spring-restrained Bellows or Bourdon tubes used to actuate conventional micro-type switches or reed-type switches. They are used where actuation at one point and low cost are the prime considerations. These instruments can be obtained to function at pressures from less than 1 to 20,000 psia. They are inherently passive devices requiring no power and little maintenance. Temperature range is -65 to $+750^{\circ}\text{F}$ (Figure 85).

2.12. TEMPERATURE TRANSDUCERS

Before selecting a temperature transducer it is important to recall a few basic characteristics of heat transfer. Heat transfers by conduction, convection, and radiation are all fundamentally different processes. Transducers that measure one process successfully are not necessarily useful for the other two. The following parameters must be weighed before selecting the instruments:

1. Temperature range of measurement.
2. Accuracy of measurement.
3. Is the material being monitored a solid or a liquid?
4. Effect of the ambient atmosphere on the measurement.
5. Readout technique to be used.

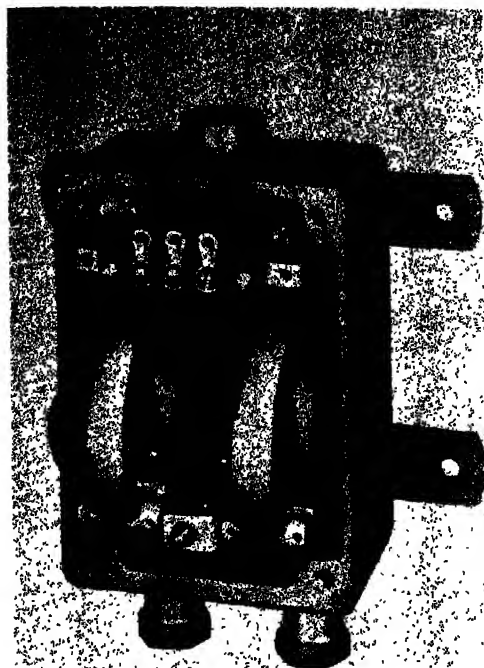


Figure 85. Bourdon-type pressure switch. (Courtesy of Meletron Corporation.)

PRELIMINARY CONSIDERATIONS. When measuring the temperature of solids, the first consideration should be the maximum allowable size of the transducer. The size of the transducer relative to the size of the test piece should be checked using the Biot modulus as a guide:

$$\text{Biot modulus} = h \left[\frac{L}{K} \right] \quad (38)$$

where h = surface heat transfer coefficient

L = smallest dimension of the solid

K = thermal conductivity of the solid

If the modulus is less than 0.2, no significant thermal gradients should exist. If the modulus is greater than 0.2, significant temperature gradients are

likely to exist in the solid, and the maximum gradient on the surface should be checked by the following equation:

$$\text{surface temperature gradient} = \frac{\Delta T}{\Delta X} = \frac{Q}{K}$$

where ΔT = change in temperature over a length ΔX

Q = heat transfer rate per unit of area at the surface

K = thermal conductivity of the solid

On the basis of the gradient, it is possible to limit the size of the temperature transducer. The largest dimension should not be greater than the distance between two points that will produce a temperature difference, due to the gradient, which will exceed the required accuracy of the measurement. Poor thermal coupling will cause a time lag that will further increase any gradient errors.

The basic problem of measuring the temperature of a liquid is to provide good thermal coupling to a substance that is inherently a poor thermal conductor. The sensing device must come to equilibrium with the temperature of the fluid before a measurement is made. If a fluid is changing temperature rapidly, the thermal capacity of the sensor must be suitable for following these fluctuations. This calculation is made as follows:

$$\frac{T_1 - T}{T_1 - T_0} = e^{-(hA/\omega C_p)\theta} \quad (39)$$

where T = temperature indicated by the sensor

T_0 = initial temperature of the sensor

T_1 = temperature of the fluid

h = heat transfer coefficient at the surface of the sensor

A = area of the sensor

ω = weight of the sensor

C_p = specific heat of the sensor

θ = time after immersing the sensor in the fluid

For efficient operation, the surface-to-mass ratio of the sensor should be large.

Another practical indicator to use is the time constant of the sensor.

$$\text{time constant} = \tau = \frac{1}{hA/\omega C_p}$$

COMMERCIAL TRANSDUCERS. Thermocouples are used for a greater variety of industrial and scientific temperature measurements than any other device. They generate their own DC output without any external excitation. Typical sensitivities are 0.005 mV/F° to 0.03 mV/F°. The output is measured on a millivolt meter, or a balancing-type potentiometer. The signal can also be used as an input to a servosystem to control temperature. Thermocouples are composed of two dissimilar metals joined at one extremity; this is normally called the hot junction. The other end of the thermocouple is maintained at a reference temperature, such as the freezing point of water. This is called the cold junction. When it is not practical to maintain the cold junction at 32°F, the actual temperature is accurately determined and an arithmetic correction is made. Standard tables are available for this purpose. The various types of thermocouples commonly used are shown in Figure 86.

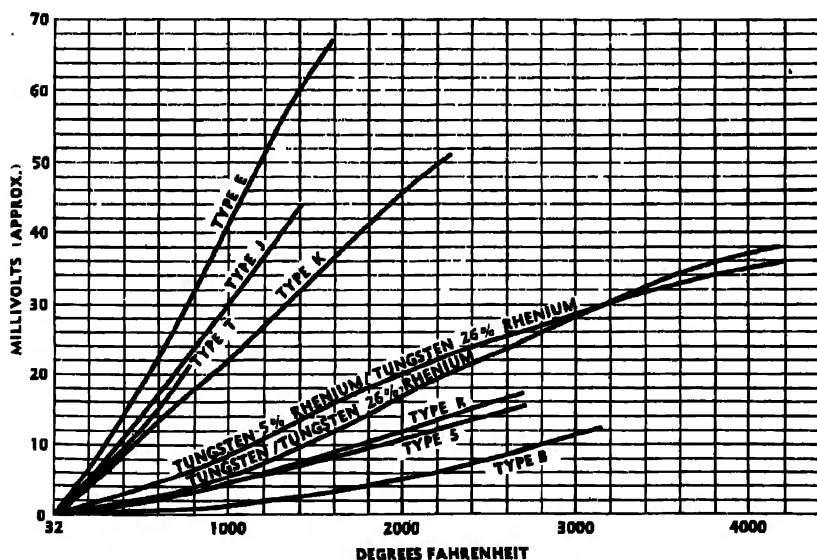


Figure 86. EMF output of typical thermocouple materials. Type E, chromel-constantan; type J, iron-constantan; type T, copper-constantan; type K, chromel-alumel; type R, platinum-13% rhodium-platinum; type S, platinum-10% rhodium-platinum; type B, platinum-30% rhodium-platinum-6% rhodium. (Courtesy of Barber-Coleman Company.)

Iron-constantan thermocouples are suitable for use in neutral or reducing atmospheres. Oxidation of the iron occurs rapidly at temperatures above



Figure 87. Resistance thermometer. (Courtesy of BLH Electronics, Inc. From Reference 3.)

1000°F. Chromel-alumel is suitable for use in oxidizing or neutral atmospheres. Copper-constantan may be used for mildly oxidizing or reducing atmospheres. It has high corrosive resistance to moisture and good repeatability. Platinum-rhodium is recommended for higher temperature application but must be protected from all types of corrosive environments to prevent contamination and subsequent calibration drift. Tungsten-rhenium is used for temperatures up to 5200°F.

RESISTANCE THERMOMETRY. This technique consists of measuring the resistance of a metal by immersing it in the environment and reading the result on a Wheatstone bridge. Although iron, copper, and other materials have been used, platinum has become the accepted standard, since it is useful over a range of -277 to $+1167^{\circ}\text{F}$ and is extremely stable. Normal accuracy is $\pm 0.02^{\circ}\text{F}$. The output may be "bucked" against a reference voltage for normal system work (Figure 87).

THERMISTORS. Thermistors are solid-state electrical elements characterized by a high negative coefficient of resistivity. Contrary to common belief, thermistors are quite stable when properly aged before use. They exhibit great temperature sensitivity and quick response time. Sensitivity is as much as ten times that of thermocouples with response times down to 0.002 sec. The practical temperature range is about -50 to $+600^{\circ}\text{F}$. Measurements to the nearest 0.1°F are possible if the current through the thermistor is limited to a value that does not increase its temperature by excessive self-heating effects (Figure 88). Commercial thermistor temperature transducers are typically $\pm 1\%$ devices.

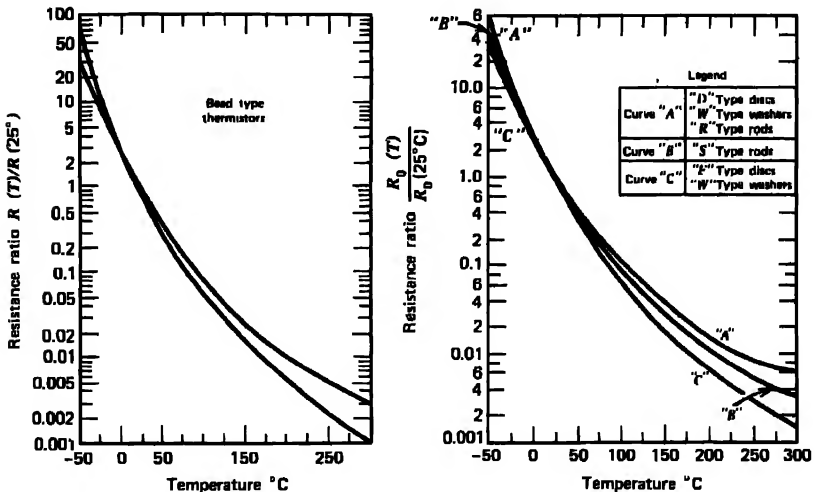
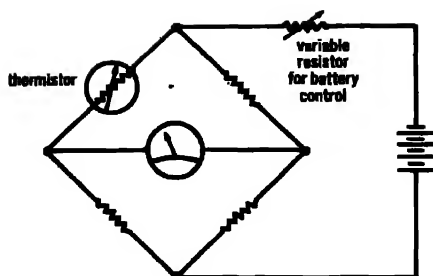


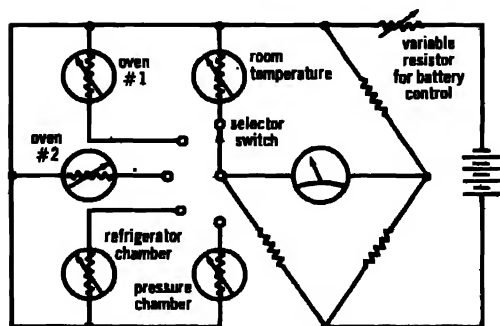
Figure 88. Thermistor temperature characteristics. (Courtesy of Benwill Publishing Company. From Reference 24.)

Some typical thermistor applications are as follows (Figure 89):

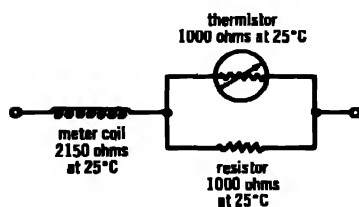
1. *Temperature measurement.* A thermistor in one leg of a Wheatstone bridge circuit will provide precise temperature information. Accuracy is limited in most applications only by the readout device (Figure 89a). Since lead length between thermistor and bridge is not a limiting factor, this basic system can be expanded to measure temperature at several locations from a central point. Thermistor interchangeability and large resistance change eliminate any significant error from switches or lead length (Figure 89b).



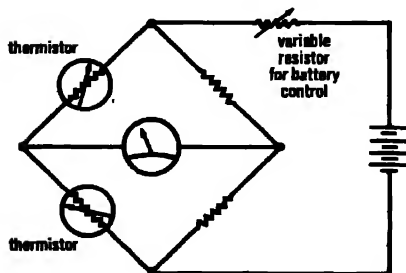
(a)



(b)



(c)



(d)

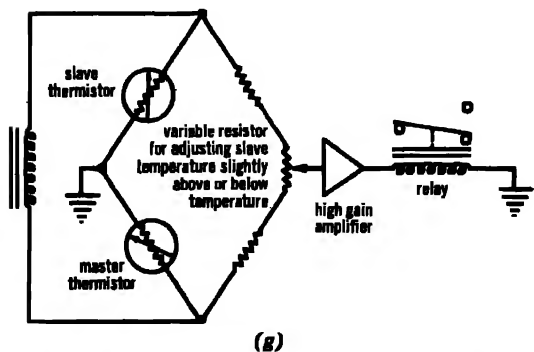
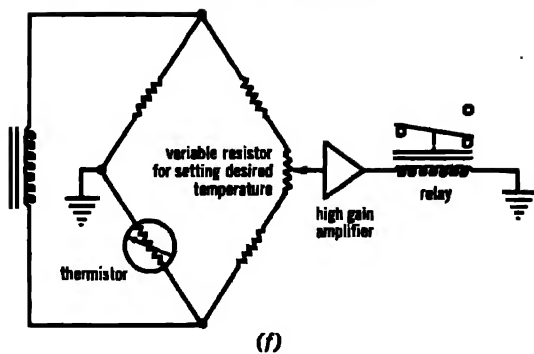
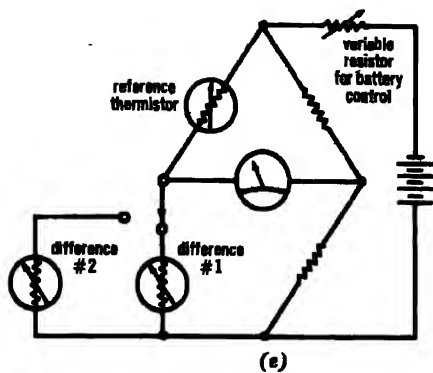


Figure 89. Thermistor sensor applications. (Courtesy of Yellow Springs Company.)

2. *Meter compensation.* The resistance of a meter movement changes with temperature, making the meter temperature dependent. Using the thermistor's property of a high negative temperature coefficient, the coil can be compensated so that total resistance due to temperature rise is essentially constant, allowing the meter to be used over a wide temperature range with minimal error (Figure 89c).

3. *Differential thermometers.* For accurate indication of temperature differential, two thermistors can be used in a Wheatstone bridge circuit. Thermistor interchangeability simplifies circuit design and reduces the number of components (Figure 89d). To measure heat loss in a piping network, thermistors can be placed at various points and the difference between these temperatures and the original temperature monitored at a convenient location. Measurement of air temperature at different elevations with reference to ground temperature is useful for temperature inversion data and geological studies (Figure 89e).

4. *Temperature control.* A system can be designed using a thermistor with a known temperature-resistance curve to form one leg of an AC bridge and a variable resistor calibrated in temperature to form another leg. When the resistor is set to a desired temperature, bridge unbalance occurs. This unbalance is fed into an amplifier which actuates a relay to provide a source of heat or cold. When the thermistor senses desired temperature, the bridge is balanced, opening the relay and turning off the heat or cold (Figure 89f).

5. *Master-slave control.* Occasionally there is a need to control one temperature with respect to another, such as a product going through a series of baths. The first bath acts as a master and uses a thermistor to sense temperature. Succeeding baths, also using thermistors, are slaves. When these thermistors are placed in the controller bridge, the slave baths can be maintained at a temperature relative to the master bath. The master bath can be controlled with the system described earlier. The master-slave controller can be used for as many baths as necessary (Figure 89g).

PYROMETERS. There are two principal types of pyrometers: optical and radiation pyrometers. An optical pyrometer is an instrument that is used to compare the brightness of an object of unknown temperature with the known brightness of a reference source of light. It measures temperature by measuring the intensity of light of a particular wavelength emitted by a hot body. The intensity of light emitted by a hot body varies greatly with its temperature. For example, the visual effect of red radiation at 2500°F varies 12 times as rapidly as its temperature. Therefore, a small change in temperature produces a great variation in brightness. By comparing the emitted intensity of an unknown with a known calibrated source, it is possible to determine temperatures with extreme accuracy. There are two

types of optical pyrometers. One compares the light from a hot body with the light from a comparison lamp in the instrument. The second type varies the intensity of light from the comparison lamp and matches its intensity to that of the unknown. An infrared filter in the instrument simulates a monochromatic condition. Pyrometers are useful up to 5200°F. They are sometimes called radiometers (Figure 90).

Radiation pyrometers differ from optical pyrometers principally in their sensitivity to radiation wavelengths. Radiation pyrometers are sensitive to all wavelengths; optical pyrometers are sensitive to only one wavelength. Radiation pyrometers use the intensity of the radiation emitted from an object as a measure of its temperature; they consist of an optical system, a temperature-sensing element, and a measuring indicator. The optical system focuses a portion of the energy radiated by the object on the sensing element. The sensor may be a thermocouple, thermopile, or thermistor. The temperature reached by the sensing element is a function of the temperature of the object observed. The pyrometer is calibrated for black-body conditions at temperatures of 1000°F or higher. Errors in the order of 20 to 30°F can be expected from nonblack-body conditions. Light-colored sources will produce even greater errors. Thermistors or thermopiles are used as sensing elements for relatively low temperatures. For temperatures above 1000°F a photoelectric or photoconductive element is used.

Black-body calibration is based on Kirchhoff's law of radiation, which states that the ratio of the energy radiated to the energy absorbed is the same for all bodies at a given temperature.

$$W_b = \frac{W}{a} \quad (40)$$

where W_b = emittance of a black body

W = energy radiated per square centimeter of surface

a = energy absorbed by the body per square centimeter of surface

The difference between reflected energy and emitted energy is important. For example, if a flashlight is directed toward a mirror, the light coming from the mirror surface is reflected energy and exhibits the characteristics of the flashlight. The light coming from the filament of the flashlight bulb is emitted energy. If the target is not a perfect black body, special calibration of the pyrometer is necessary. If the target is transparent or translucent, the pyrometer may also be "looking" at the surface behind it. One approach to this problem is to use a light source operating at a frequency that is not transmitted by the target. The background receives no radiation and cannot influence the measurement. Under ideal conditions, pyrometer accuracy is about $\pm 0.5^\circ\text{C}$. If the conditions are not ideal, errors may approach 50%.

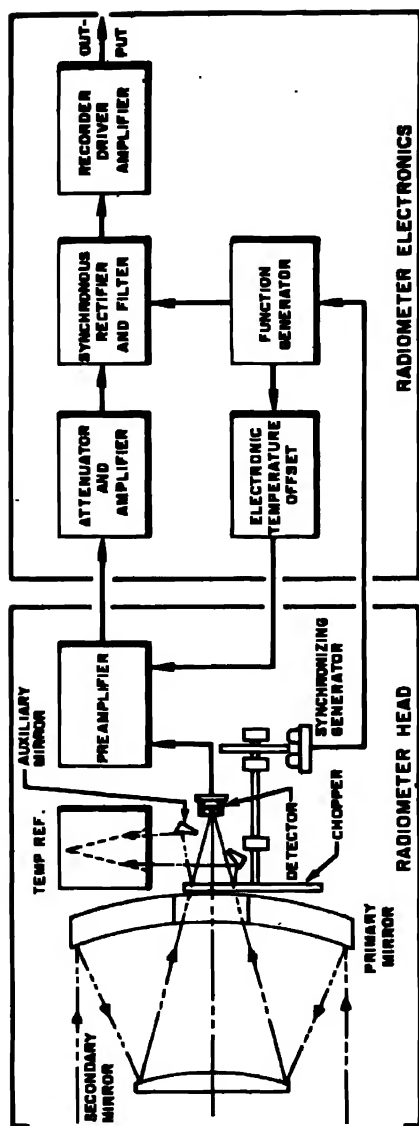


Figure 90. Radiometer optical schematic and electronic block diagram. (Courtesy of Barnes Engineering Company. From Reference 23.)

REFERENCES

1. *Electromechanical Design*, January 1967, Benwill Publishing Company, Brookline, Mass.
2. *Helipot Components Catalog B-27*, Helipot Division, Beckman Instrument, Inc., Fullerton, Calif., 1968.
3. *Foil Strain Gage Specifications*, BLH Electronics, Inc., Waltham, Mass., August 1967.
4. P. R. Perino, *Characteristics of Thin Film Strain Gage Transducers*, Instrument Society of America, Pittsburg, Pa. 1966.
5. *Thin Film Strain Gage Pressure Transducers*, Statham Corporation, Oxnard, Calif., 1962.
6. *Semiconductor Strain Gages*, Kulite Corporation, Ridgefield, N.J., n.d.
7. C. M. Harris and C. E. Crede, *Shock and Vibration Handbook*, Vol. 1, Chapter 17, McGraw-Hill Book Company, New York, 1968.
8. *Notes on Linear Variable Differential Transformers*, Schaevitz Engineering Company, Pennsauken, N.J., 1955.
9. *Intrinsic Sensor FR-1*, M.T.I. Corporation, Latham, N.Y., July 1967.
10. Capacitance Transducers, *Product Engineering*, November 7, 1960, McGraw-Hill Book Company, New York.
11. E. Charkey, Measuring Without Touching, *Electromechanical Design*, November 1966, Benwill Publishing Company, Brookline, Mass.
12. J. T. LaVan, Resolvers and Synchros, *Electromechanical Design*, January 1965, Benwill Publishing Company, Brookline, Mass.
13. R. M. Kerchner and G. F. Corcoran, *Alternating Current Circuits*, John Wiley & Sons, Inc., New York, 1960.
14. *Technical Information for the Engineer*, Singer-General Precision, Inc., Little Falls, N.J., 1968.
15. *Gimbal Resolvers/Synchros*, Reeves Instrument Company, Garden City, N.Y., 1964.
16. *Synchros-Motors-Rotating Components Catalog*, Clifton Division, Litton Industries, Drexel Hill, Pa., 1968.
17. V. Del Toro and S. R. Parker, *Principles of Control System Engineering*, McGraw-Hill Book Company, New York, 1960.
18. H. Chestnut and R. W. Mayer, *Servomechanism and Regulating System Design*, 2nd ed., John Wiley & Sons, Inc., New York, 1959.
19. *Direct Drive Servo Design*, Inland Motor Division, Rollmorgen Corporation, Radford, Va., 1965.
20. *Dynamic Instrumentation for Shock, Vibration, Pressure, Displacement*, Columbia Research Laboratories, Inc., Woodlyn, Pa., 1969.
21. J. P. DenHartog, *Mechanical Vibrations*, McGraw-Hill Book Company, New York, 1965.
22. *Series 2200 Accelerometer Manual*, Endevco Corporation, Pasadena, Calif., 1964.
23. *Theory, Application, and Instrumentation for Infra-Red Non-Destructive Testing*, Barnes Engineering Company, Stamford, Conn., 1964.
24. *Electromechanical Design*, July 1965, Benwill Publishing Company, Brookline, Mass.

Chapter III Encoders

Encoders are the primary instruments for converting analog information to digital information. Any physical process, displacement, velocity, or acceleration that is normally monitored by the rotation of a dial can be converted to digital form. The broadest definition of an encoder includes all devices that transform analog information into a format useful to a readout device. For example, the circuitry in a digital voltmeter converts the voltage to be measured to a form that can be interpreted by a decade counter. A mechanical encoder is exemplified by an auto odometer. However, by convention an encoder is more commonly thought of as an instrument that converts angular rotation of a shaft into a train of pulses. The pulses generated are in accordance with a particular code that is compatible with the readout instrument.

Encoders are used when system accuracies require a digital, rather than an analog approach. As explained in Chapter 1, the inherent error of a digital system is approximately ± 1 count per component. The usual accuracy of analog devices is $\pm 0.1\%$ at best, and more commonly ± 2 to 5% . Comparable accuracy of an encoder ranges from $\pm 0.1\%$ at worst to $\pm 0.0001\%$ for very good units.

3.1. CONSTRUCTION

The three primary types of encoders are brush, optical, and magnetic. Each type contains a disk, mounted between bearings, that is mechanically coupled to the analog-sensing mechanism. The disk contains some form of coded pattern that is used to make and break a circuit. The method of making and breaking the circuit is the basis of the three primary classifications.

3.1.1. Brush Encoders

Brush encoders consist of two main components: the encoding disk and the brush assembly. The disk contains a series of concentric rings or tracks that are photographically printed on its surface (Figure 1). Commercial

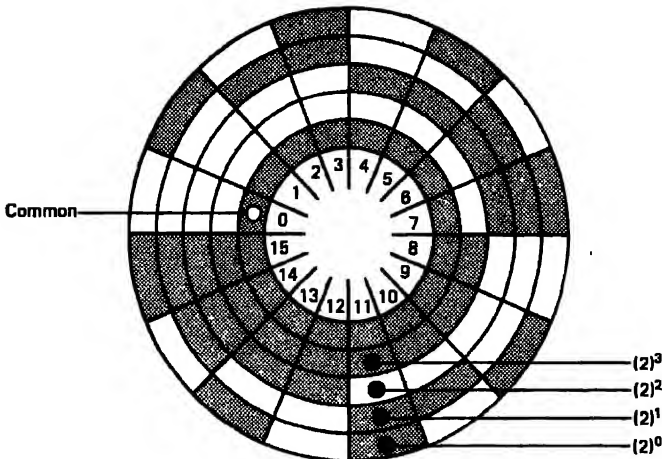


Figure 1. Binary code disk. (Courtesy of Litton Precision Products.)

encoders contain 2 to 30 tracks; one is continuous for 360° and the rest are segmented. The brush block assembly contains at least one brush for each track. The continuous track is the "common." The segmented tracks are separated by insulating material, and individual segments in each ring are electrically shorted together. As the encoder disk rotates, the common brush and certain of the others periodically form a closed circuit by contacting conductive strips and then open the circuit as one brush rests on an insulator. The result is a square-wave pattern that can be used to identify the position of the disk. The disk is usually made from a printed circuit board with suitable mechanical backing. Any pattern that can be produced photographically can be reproduced on the disk. Encoders commonly utilize uniformly spaced tracks, nonuniform tracks, adjacent tracks at some fixed phase angle, and numerous other unusual configurations for exotic system control functions. However, the most common application, measuring the shaft position, requires tracks that are uniformly spaced.

For example, consider an encoder disk that contains two rings, one continuous and one broken into 1024 conductive segments separated by

1024 insulating segments. On the back side of the disk, all 1024 conductive segments and the continuous ring are shorted together. Every time the disk rotates through $2 \times 1/1024$ of a revolution, a complete square-wave pulse is produced. This makes it possible to measure the position of the disk to the nearest $1/1024$ of a revolution or 21 min of arc. Suppose the two brushes are wired in series with a battery and an electronic counter. For every revolution of the shaft, the counter will read 1024. If it is desirable to key the counting process to one starting point on the disk, it is necessary to reset the counter after the 1024th square wave is recorded. This can be accomplished by providing a third track that produces a reset pulse just after the 1024th pulse and prior to the first pulse. The reset pulse is directed to the electronic counter, and every cycle will start with the counter set at zero. If additional tracks were provided, many other types of synchronizing and control pulses could be developed. The encoder used in this example is usually referred to as a 10-bit encoder, since it provides 2^{10} or 1024 bits per revolution of the disk.

Two basic errors are associated with encoders used for measuring shaft position: nonuniform pattern spacing and eccentricity. Nonuniform spacing is caused by inaccuracies in manufacturing the master pattern from which all successive encoders are made. The pattern is made on an indexing table that is similar to those used on milling machines, but is much larger and more accurate. For small commercial encoders, this error is normally 5 to 10 sec of arc. Expensive, military units are often made to ± 0.1 sec. The net result of this defect is that the pulses formed by the encoders are irregular. The second error, pattern eccentricity, is more subtle but produces a once-per-revolution error (Figure 2). If the center of the pattern is eccentric with respect to the center of rotation of the disk, the velocity of the track will change in proportion to its radius. During every revolution there is a sinusoidal change in position that will be reflected as a velocity error in the encoder output. One method of minimizing this problem is to use two brushes per track, spaced 180° apart. When one brush is at a minimum velocity position, the other is at a maximum; if the two pulses are added electronically, the net result is almost independent of eccentricity errors. This error is analogous to listening to a phonograph record with "wow." A sustained note will have a variation in pitch due to eccentricity of the groove in the record or changes in the disk velocity.

Brush block assemblies are designed to produce line contact between the brushes and the segments of a given pattern; in this way, the width of the segments can be minimized. Practical limitations make this about 0.005 to 0.010 sec. The brush pressure must be adjusted so that it is light enough not to cause excessive disk wear but heavy enough not to suffer from high contact resistance or bouncing effects. When the number of segments

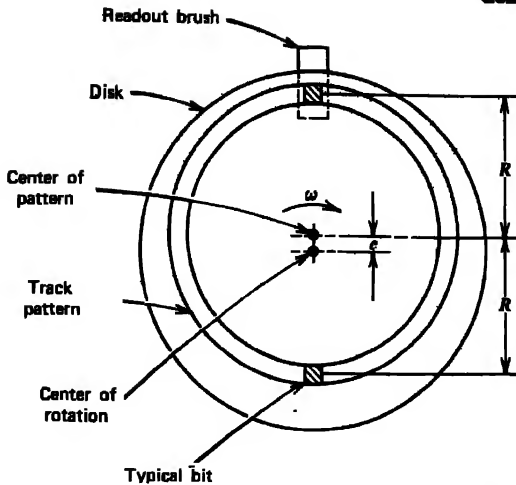


Figure 2. Code disk eccentricity error. Pattern eccentricity, e , changes velocity of bits as they pass readout brush. A typical bit may have a velocity of $\omega(R + e)$; another bit 180° away will indicate $\omega(R - e)$ as it sweeps past the brush. Intermediate bits will have proportionally varying velocities. (Courtesy of Benwill Publishing Company. From Reference 1.)

contacted per second is high, the natural frequency of the brush assembly must be considered or brush resonance may result.

Brush encoders are used wherever modest budgets and medium accuracies are required. This includes the vast majority of all industrial applications. Typical instruments are 10- to 12-bit units. Since they are the oldest type of encoder, they are usually thoroughly "debugged." In general, the accuracy of a large-size encoder is better than a small encoder for a fixed amount of money.

3.1.2. Optical Encoders

Optical encoders were developed to meet the requirements of space-age technology. They are not limited by the speed and wear considerations inherent in brush encoders. Nineteen-bit optical encoders are considered standard hardware today. Optical encoders are composed of three primary parts: a rotating disk, a light source, and a light detector (Figure 3). The tracks on the disk are composed of transparent and opaque patterns produced by exposing a photographic emulsion to light. The emulsion is mounted on a glass disk to provide mechanical rigidity and dimensional stability. The bit width and spacing are normally less than one tenth the

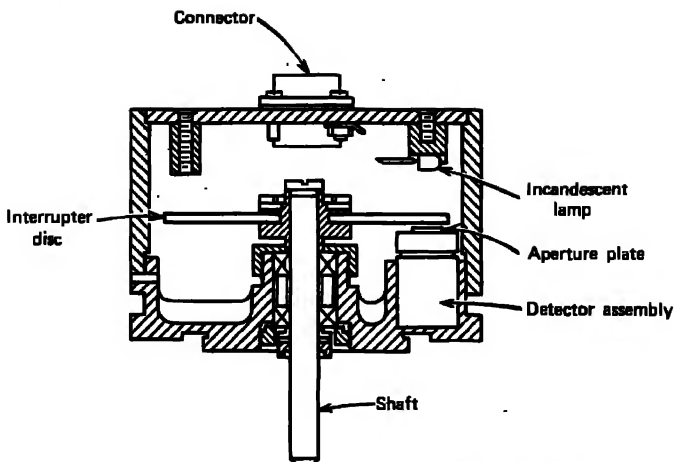


Figure 3. Optical encoders. (Courtesy of W. & L. E. Gurley Corporation.)

size of brush encoder patterns. Segments a few ten thousandths of an inch in width are in use in commercial encoders. The light source is a subminiature tungsten bulb or a solid-state device. The detector is usually a miniature solid-state device selected for optimum sensitivity to the radiation produced by the source. The light source is optically focused on the plane of the rotating disk to use the available light most efficiently. This also increases the signal-to-noise ratio of the instrument.

Optical encoders are used whenever the utmost accuracy is required and price is secondary. Units with 21-bit accuracy are available when required. The aerospace industry has been the chief catalyst in developing these instruments.

3.1.3. Magnetic Encoders

Magnetic encoders are used when moderate accuracy and high encoder shaft speeds are required. Typical instruments have 8-bit capacities and can be used at speeds up to 10,000 rpm. They represent a stage in encoder technology midway between the brush and optical instruments. Magnetic encoders provide the inherent reliability and low life of nontouching technology, but not the resolution of the newer optical techniques.

Operation of the encoder is based on a magnetic saturation principle (Figure 4). A disk containing ferritic material is magnetized so that it has a flux pattern representing a particular track. Each track is read by "square

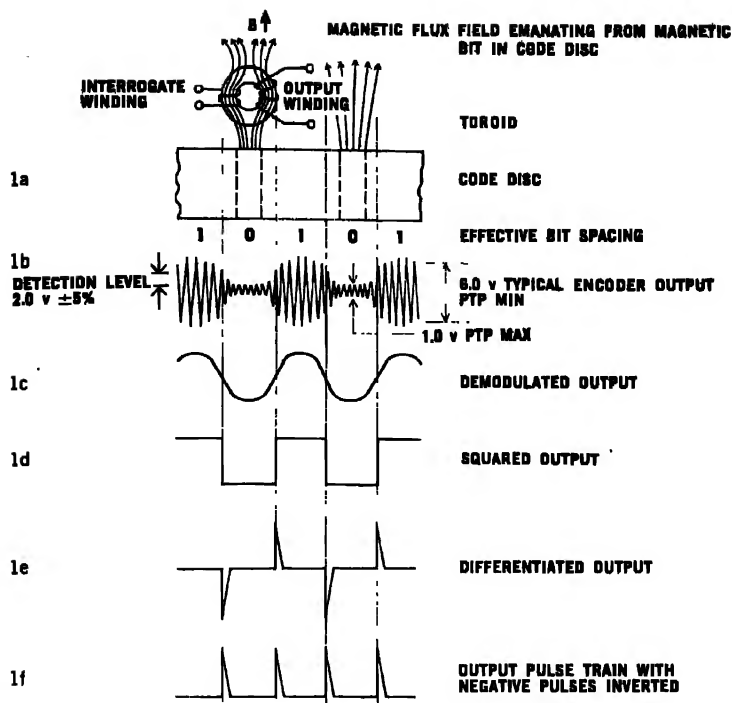


Figure 4. Magnetic encoder—principle of operation. (Courtesy of Norden Corporation.)

loop" toroids that sense the code on the disk by the difference in impedance; this variation is produced by the presence or absence of a magnetic field through the toroid. Each toroid has two windings, one for excitation (interrogation) and one for readout (output winding). Usually the excitation windings of all toroids in an encoder are wired in series and are excited at a frequency of 20 to 200 kHz. The higher the frequency, the better the resolution. When a toroid is over a magnetic area on the disk, it is saturated, and very little output is obtained from the secondary winding. When the toroid is over a nonmagnetized area, it is not saturated, and a high output appears at the secondary. The output from each track sensor is therefore an amplitude-modulated signal in which the peaks represent logical "1"s on the disk and the valleys represent logical "0"s. This amplitude-modulated signal is then demodulated, level-detected, and squared by auxiliary circuitry to provide a DC representation of the shaft position. It is next differentiated and either the positive or negative pulses are inverted to

provide the proper polarity pulses to work into the associated data-handling logic circuits.

The encoder may be interrogated continuously or on command. The input excitation may be sine wave, square wave, or intermittent pulses, so long as it consists of alternating positive and negative excursions. The amplitude and rate of change of the excitation must also be adequate to switch the polarity of the reading head.

3.2. ABSOLUTE VERSUS INCREMENTAL ENCODERS

The three types of encoders can be further subdivided into two primary categories: absolute and incremental encoders. Absolute encoders provide data that indicate some unique angular position. Incremental encoders provide data that indicate the angular excursion past some fixed reference point.

Absolute encoders contain a disk with multiple track patterns. For a given angular position, a unique combination of coded information is presented by the encoder (Figure 1). One "bit worth" of a revolution away, another unique combination is read. More details on the codes used are given in Section 3.2.1. The combination of information presented in each angular position identifies it as effectively as a number identifies a house. The decoding operation is performed by solid-state circuitry and presented in visual form, such as nixie tubes.*

Incremental encoders can have only one track pattern. Each track contains a series of identical, uniformly spaced marks. The significance of each mark is simply its distance from the starting point. The starting point is previously established by mechanical design. As the disk rotates, each subsequent pulse is presented to an electronic counter and then visually displayed. No coding system is necessary. The net result of this system is a very precise measurement of angular motion relative to a fixed point. If the output is differentiated with respect to time, it produces an excellent tachometer.

Modern systems tend to use more incremental encoders than absolute encoders for the following reasons. Incremental encoders use fewer tracks than absolute encoders; consequently, the number of leads, slip rings, readout devices, circuitry, and display elements are kept to a minimum. This produces greater reliability and lower costs. The basic disadvantage is that measurements are only relative to one fixed point. If that point is in

* The term nixie tubes is used here in the generic sense. NIXIE TUBES are a product of the Burroughs Corporation.

error, the whole system suffers. Another problem is that if a power failure occurs the data are usually lost. This can be prevented if auxiliary data-storage techniques are used.

The absolute encoder is utilized where absolute positions must be known at all times even if power is interrupted for a year. It is also less affected by noise and extra pulses that are parasitically produced by other equipment.

3.2.1. Coding

When absolute encoders are required some form of coding system must be used to uniquely define each position. Our conventional numeral system is based on powers of 10. For example, the number 7,367.8 means

$$7 \times 10^3 + 3 \times 10^2 + 6 \times 10^1 + 7 \times 10^0 + 8 \times 10^{-1}$$

Practical digital circuitry is based on powers of 2. The binary coded number 1001 means

$$\begin{aligned} 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ = 1 \times 8 + 0 \times 4 + 0 \times 2 + 1 \times 1 = 9 \end{aligned}$$

The zero and 1 represent two opposite electrical states. This may be a conducting and nonconducting state, or a wide difference in voltage levels. The important consideration is that conventional digital circuitry can utilize the system for high-speed counting applications.

Suppose that we wanted to represent the number 5280 in binary form. It would be

$$\begin{aligned} 1 \times 2^{12} + 0 \times 2^{11} + 1 \times 2^{10} + 0 \times 2^9 + 0 \times 2^8 + 1 \times 2^7 \\ + 0 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 \\ + 0 \times 2^2 + 0 \times 2^1 + 0 \times 2^0 \\ = 4096 + 0 + 1024 + 0 + 0 + 128 + 0 \\ + 32 + 0 + 0 + 0 + 0 + 0 \\ = 5280 \end{aligned}$$

This could be written as 101001010000.

To avoid some of the problems inherent in binary formats, the binary coded decimal system, BCD, was created. Each power of 10 is assigned 4 binary digits. The number 5280 is represented as follows:

$$\begin{array}{cccc} \underline{0101} & \underline{0010} & \underline{1000} & \underline{0000} \\ 5 & 2 & 8 & 0 \end{array}$$

After some experimenting with binary and BCD systems, it became apparent that if, due to some mechanical problem, a brush contacting one row of digits accidentally touched an adjacent row, a serious error could result. To eliminate this possibility, the Gray code was developed. The Gray code is obtained from the binary code in the following way. Proceeding from left to right, copy the original numeral but change any digit immediately preceded by 1 to its opposite state (Figure 5). For example, 1110 binary equals 1001 in Gray code.

Decimal	Natural Binary	Cyclic Binary (Gray Code)	Excess 3	Binary	Decimal	Cyclic Binary Decimal
0	0000	0000	0011	0000	0000	0000 0000
1	0001	0001	0100	0000	0001	0000 0001
2	0010	0011	0101	0000	0010	0000 0011
3	0011	0010	0110	0000	0011	0000 0010
4	0100	0110	0111	0000	0100	0000 0110
5	0101	0111	1000	0000	0101	0000 0111
6	0110	0101	1001	0000	0110	0000 0101
7	0111	0100	1010	0000	0111	0000 0100
8	1000	1100	1011	0000	1000	0000 1100
9	1001	1101	1100	0000	1001	0000 1000
10	1010	1111	1101	0001	0000	0001 1000
11	1011	1110	1110	0001	0001	0001 1001
12	1100	1010	1111	0001	0010	0001 1011
13	1101	1011	10000	0001	0011	0001 1010
14	1110	1001	10001	0001	0100	0001 1110
15	1111	1000	10010	0001	0101	0001 1111
16	10000	11000	10011	0001	0110	0001 1101
17	10001	11001	10100	0001	0111	0001 1100

Figure 5. Comparison of various code systems. (Courtesy of Benwill Publishing Company. From Reference 1.)

The most important reason for the use of Gray code is the elimination of ambiguity between successive rows of digits. This is discussed in detail in section 3.3.1.

Cyclic binary decimal combines the BCD and Gray codes. For example, 5280 in cyclic binary decimal is represented as

$$\begin{array}{ccc} \underbrace{0111} & 1100 & 0000 \\ 5 & 8 & 0 \end{array}$$

Another code in use is the excess three. It is similar to Binary with all the numerals displaced downward by three places. To convert from the binary code to excess three, add 11 to each number. For example,

$$\begin{array}{rcl} \text{binary code 6} & = & 0110 \\ & + & 11 \\ \hline \text{excess three code 6} & = & 1001 \end{array}$$

This system is not extensively used today. There are numerous other codes in use, each with some small computational advantage for a particular problem. To summarize, use the binary code where the output of the encoder is supplied directly to a computer. When the output is to be viewed on nixie tubes, use BCD. For applications where the encoder output is fed to other data-processing equipment with a particular code, use the same code. In most cases, this is the Gray code. If you were designing a system that would count from 0 to 128, the number of tracks necessary on the encoder would be as follows:

Binary: 8

BCD: 12.

Cyclic binary decimal: 12.

Gray code: 8.

3.3. DESIGN CONSIDERATIONS

As previously indicated, one of the most important design goals is to have as many bits per track as possible; in this way the accuracy of the instrument is increased. One method of increasing accuracy is simply to increase the size of the encoder. It is easier to get more bits on the periphery of a large disk than on a small one. This approach obviously has limitations. Most encoder envelopes are between 1 and 2 in. in diameter. The upper range is usually about a 4 in. diameter. The practical limitations are rotor inertia, bearing run-out, shock, and vibration effects. A standard practice has been to install a gear train between the encoder disk and subsequent "fine reading" disks. In this way, finer increments can be read in much the same fashion as a second hand is read on a clock. Unfortunately, this is no panacea, because the inherent accuracy on the fine increments is no better than that of the first disk. Furthermore, it is decreased by the inherent errors in the gear train.

3.3.1. Least and Most Significant Digit

The important design parameters are the least significant digit, LSD, and the most significant digit, MSD. The least significant digit is located on the outmost track and is defined as follows:

$$\text{arc subtended by LSD} = \frac{360^\circ}{2^N}$$

where N = total number of binary digits on the track. The most significant digit is located on the innermost track and is defined by

$$\text{arc subtended by MSD} = 2^{N-1} \times \frac{360^\circ}{2^N}$$

In Figure 1, the LSD is read by the 2^0 brush and the MSD is read by the 2^3 brush.

3.3.2. Ambiguity

A chronic problem of multibrush encoders is readout ambiguity (Figure 6). A line of brushes is illustrated at three different positions. Consider the

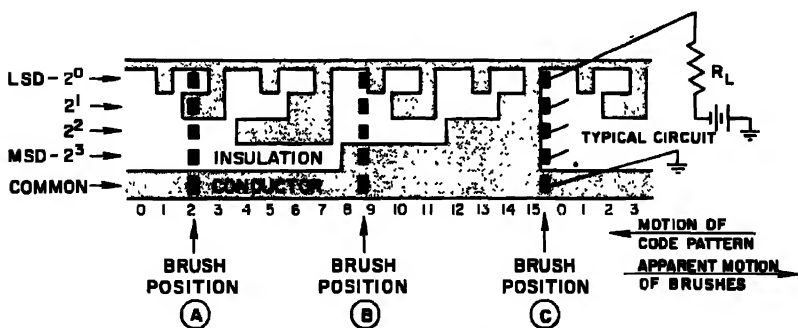


Figure 6. Definition of encoder ambiguity. (Courtesy of Singer-General Precision, Inc. From Reference 3.)

brushes to be stationary and the code pattern to be moving from right to left. At A, the brushes read binary 0010. (The MSD brush is read first and the LSD last.) The reading is without error. At B, the brushes are close to the halfway point between 1000 and 1001. Either number may be indicated, depending whether the 2^0 brush is conducting or not. Either reading is

equally correct and within the permissible error of ± 1 count. At *C*, the brushes can read any of the 16 possible code combinations due to minor code pattern boundary irregularities and finite brush misalignments. This condition illustrates ambiguity and can be eliminated by using the Gray code (Figure 7). This system has conductive strips that overlap so that no

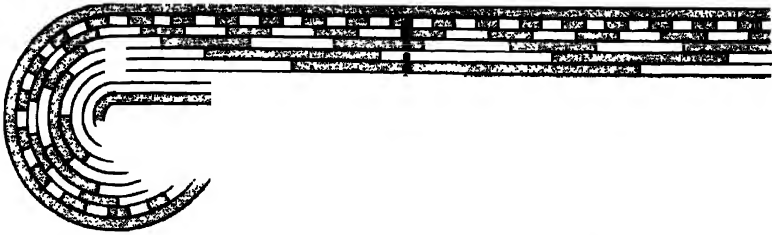


Figure 7. Gray code encoder. (Courtesy of Singer-General Precision, Inc. From Reference 3.)

two or more brushes can "change state" simultaneously. Any minor brush misalignment is incapable of causing a position error in excess of 1 digit. The encoder is therefore accurate to ± 1 count. Note that, although the term "brush" is used in this section, the principles apply equally to optical and magnetic encoders.

The V-Scan encoder is another method of eliminating ambiguity. When a zero (nonconducting area) appears on any track at any position, the next higher order digit-track has an area that does not change state in the forward direction for double the digit width (Figure 8). Conversely, when



Figure 8. V-scan encoder. (Courtesy of Singer-General Precision, Inc. From Reference 3.)

a "1" (conducting area) appears on any track, the next higher-order digit has not changed state in the backward direction for double the digit width. A single brush is used on the least significant digit track. Two brushes are

required on each higher-order track—one leading the reading line (advanced brush) and one lagging the reading line (retarded brush); they are switched logically in and out of the readout by auxiliary circuitry to avoid reading a brush as it switches at a boundary between code counts. The logic circuitry for brush switching may be packaged separately or as a module attached to the encoder. The rules of logic are as follows: when a brush reads “1,” go to the next higher-order retarded brush. If it reads “0,” go to the next higher order advanced brush. V-Scan is particularly useful when the encoder is used for direct computer input.

Another method of eliminating ambiguity is the U-Scan encoder (Figure 9). The least significant digit acts as a selector bit that causes the drive

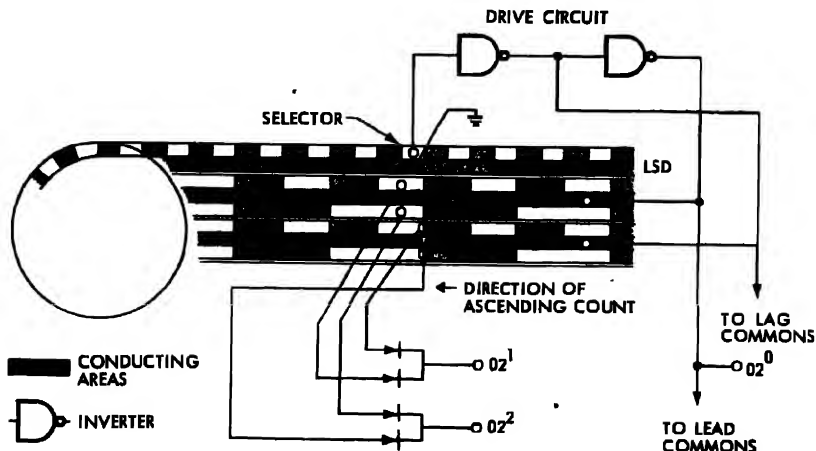


Figure 9. U-scan encoder. (Courtesy of Singer-General Precision, Inc. From Reference 3.)

circuit to energize either the lagging or the leading brushes. Power is switched so that no brushes are ever energized when crossing from a conductive to an insulated area, or vice versa. This eliminates ambiguity by allowing simultaneous readings to be made at one preselected instant. It also eliminates wear associated with commutation of current across switching points. The name of this configuration is derived from the position of brushes that resemble an inverted U. The rules of logic for the drive circuit are as follows. When the selector bit is on, energize the lagging brushes. When the selector bit is off, energize the leading brushes. The purpose of the diodes is to block transients that would activate circuit loops.

3.3.3. Direction of Rotation

With many system problems, it is important to know the direction of rotation of an encoder at any instant. A change in direction could indicate a subtractive process, a change in velocity, a variation in quantity, or many other significant events. The method described in this chapter was developed primarily for magnetic encoders but, with slight modification, can be used in other types of encoders. The technique requires two brushes or coils monitoring an LSD track. The brushes are spaced 90 electrical degrees apart with respect to the square wave produced by an LSD track (Figure 10). Let A represent the output of the first (counting) brush and B the output of the second (direction-sensing) brush. Let DA represent the differentiated transition from the logical state "0" to the logical state "1" of output A ,

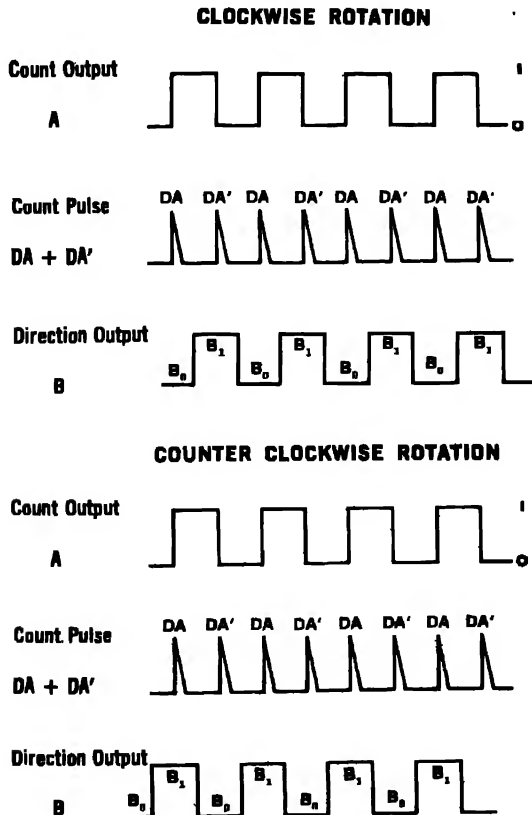


Figure 10. Sensing encoder direction of rotation. (Courtesy of Norden Corporation.)

and DA' the differentiated and inverted transition from logical state "1" to the logical state "0" of output A .

During clockwise rotation of the encoder, a DA pulse occurs only when the direction-sensing brush reads logical "0" (B_0), and a DA' pulse occurs only when the direction sensing brush reads logical "1" (B_1).

A similar analysis for the counterclockwise rotation shows that a DA pulse occurs only when the direction-sensing brush reads logical "1" and a DA' pulse occurs only when the direction-sensing brush reads logical "0." To understand the reason for this derivation, remember that when the encoder rotates clockwise, the direction-sensing brush *lags* the counting brush by 90° ; when the rotation is counterclockwise, the direction-sensing brush *leads* the counting brush 90° . The logical schematic is shown in Figure 11.

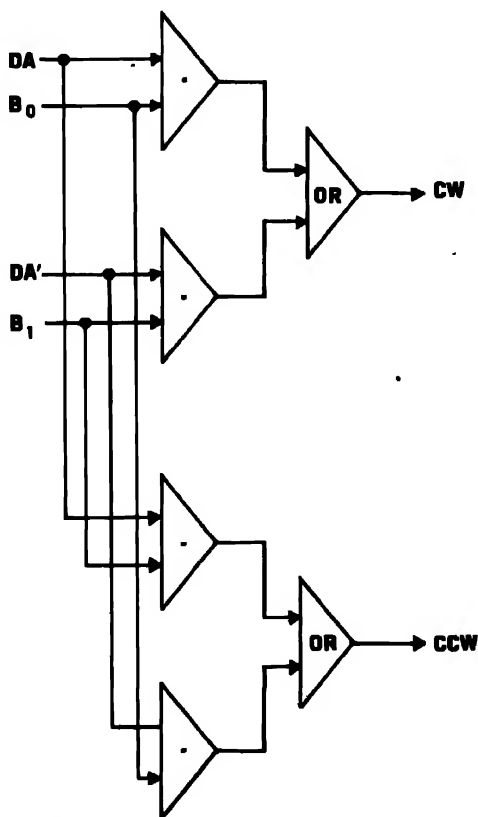


Figure 11. Direction-sensing logic circuitry. (Courtesy of Norden Corporation.)

3.3.4. Hardware Problems

In addition to the profound system problems inherent in encoder usage, there are also a few chronic, mundane hardware problems that must be solved before anything will function.

3.3.5. Bearings

Poorly selected bearings can cause once-per-revolution errors, vibration that can obscure electrical readout data, and lower encoder life. The easiest rule of thumb is that the lowest quality bearing that should be found in an encoder is ABEC 4. Good encoders should be equipped with ABEC 5 or 6. The best units are built with ABEC 7 or special bearings.

Errors due to bearing run-out and eccentricity can be minimized by mounting the encoder disk between its own bearings. In this way the center of rotation of the disk can be adjusted to almost coincide with the center of the supporting shaft. Coupling the encoder shaft to the shaft being monitored presents another problem, since the eccentricity, or run-out of the two shafts with respect to each other causes cyclic errors. Special coupling shafts for encoders are available to minimize this problem.

When buying an encoder, look at the specifications on starting torque required and the moment of inertia about the output shaft. Check to see if they are compatible with your system. They are also valuable criteria on the soundness of the mechanical design of the encoder.

3.3.6. Light Sources

The selection of a light source is, of course, a problem limited to optical encoders. Two primary types are currently available: tungsten and solid-state sources. Tungsten bulbs have been used in encoders for about 20 years. Much is known about how to make them and how to prolong their lives. Filaments are tiny, tightly wound helical elements with a high natural frequency. The chief source of failure is the accidental application of excessive excitation voltage. Typical lamp life is 5000 hours, but this can be extended by a factor of 3.5 when the nominal voltage is decreased by only 10%. Solid-state light sources generate only 0.1 to 0.01 the illumination produced by comparable tungsten sources, but they are smaller and have a much higher resistance to shock and vibration environments. Gallium arsenide semiconductors are the principal solid-state light sources used today. They are available in packages less than $\frac{1}{8}$ in. in diameter and $\frac{1}{4}$ in. high and are equipped with an integral lens. Typical emission spectra peak at about 6500 angstrom units (\AA). The associated sensor is usually a photoconductive or photovoltaic light sensor that has compatible sensitivities. Other solid-state systems operate at about 9000 \AA and are totally invisible. They are

therefore less affected by ambient light conditions. The cost of these units is higher than equivalent tungsten-type encoders. Another objection is the basic lack of reliability data at the present time. This objection should be eliminated within 2 years.

3.4. READOUT DEVICES

Most systems using encoders require a visual readout of the data developed by the instrument. Typical applications include radar antenna position indicators, airborne data acquisition, machine tool positions, plotting tables, liquid level measurements, and automatic weight logging. The readout device is usually some form of visual display tube equipped with appropriate logic circuitry. Absolute encoders require binary-to-decimal logic circuitry so that the instantaneous position of the device is constantly available.

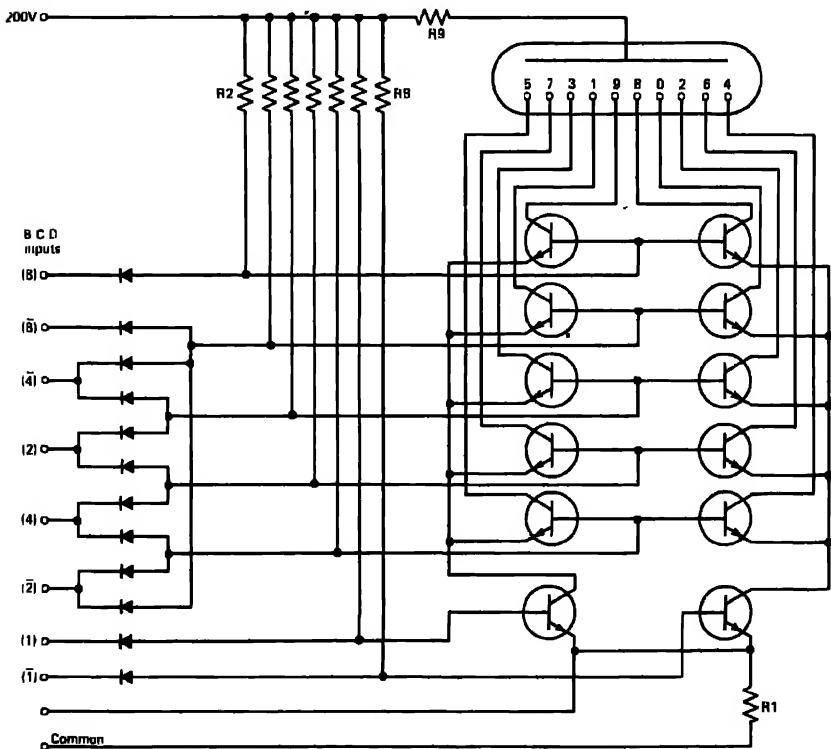


Figure 12a. Decoder driver circuit. (Courtesy of National Electronics Corporation.)

Incremental encoders simply require a very high-speed pulse counter that may be reset periodically.

3.4.1. Circuitry for Absolute Encoders

The most common figuration in use is a BCD encoder that is coupled to a decimal readout. The logic circuitry used is a basic "AND" circuit. It requires 8 logical inputs plus a common input (Figure 12). This would

BCD Input	Displayed Character									
	0	1	2	3	4	5	6	7	8	9
1	0	1	0	1	0	1	0	1	0	1
$\bar{1}$	1	0	1	0	1	0	1	0	1	0
2	0	0	1	1	0	0	1	1	0	0
$\bar{2}$	1	1	0	0	1	1	0	0	1	1
4	0	0	0	0	1	1	1	1	0	0
$\bar{4}$	1	1	1	1	0	0	0	0	1	1
8	0	0	0	0	0	0	0	0	1	1
$\bar{8}$	1	1	1	1	1	1	1	1	0	0

Figure 12b. Decoder driver circuit truth table. To display a 7, the required inputs are 1, 2, 4, 8. (Courtesy of National Electronics Corporation.)

be sufficient to program a single tube for values between 0 and 9. If we required a system with a readout from 0 to 99, two decades and two tubes would be required. For a system up to 999, three decades are required. Each tube has its own set of logic circuitry. It is generally referred to as a "nixie" driver circuit. Obviously this involves a lot of circuitry for a multidecade system. Sometimes it is called parallel word output. If we could sample each four track group on the encoder, send it to the appropriate tube, and then switch to the next set of tracks with its associated tube, we would save a lot of circuitry. This is often done and is called serial word output. One set of logic circuitry is shared for the total operation. The switching operation can be performed by a variety of auxiliary circuitry. The price exacted by this saving in circuitry is slightly reduced response time. Normally this is no problem, since the response of the circuit is many times faster than the human eye can follow. For example, suppose that we have position 4567 on an encoder. The least significant 4-bit digit is read first; then the second,

third, and fourth. Each digit has its own common track and uses the same logic circuitry to present its digit in parallel form. The first tube would receive the inputs 0111 (7), the second, 0110 (6), the third, 1000 (5), and the fourth, 0100 (4). If a parallel word system is required, four identical sets of logic circuitry would be required. The circuit in Figure 12 would also require logical inputs for the 1, 2, 4, 8 inputs.

If plus and minus BCD encoders are used, additional circuitry is required. This might be the case for a $\pm 180^\circ$ -system. Standard logic packages on plug-in cards are available.

So far we have discussed only BCD-to-decimal conversion circuits. Also available are Gray-to-decimal and excess-three-to-decimal logic circuits as well as many other special combinations of decoding systems.

3.4.2. Circuitry for Incremental Encoders

The circuitry used for incremental encoders is fundamentally simpler than that used for absolute encoders. The pulses produced are simply totaled by an electronic counter. The number of counts per revolution or cycle is predetermined, and a simple relationship exists between the total count and the position of the encoder. Often it is desirable to reset the counter every revolution. This can be accomplished by a second track on the encoder disk. The counter must be capable of subtracting as well as adding to properly accommodate changes in direction. The only problem in this system is that in case of a power interruption the count is interrupted and the reading may be worthless. It is also more susceptible to noise or false pulses induced by nearby auxiliary equipment. Normally these problems are eliminated by carefully shielding all leads.

Although we have concentrated on visual displays of encoder outputs, numerous other techniques are employed. Data storage on magnetic type and high-speed strip chart recorders are common. The response time of the recording medium is often a limitation on the system, since logic decoding is normally measured in nanoseconds while "high-speed" recording is many times slower.

A very useful *variation in visual display* tubes is the rear-projection readout technique (Figure 13). This method makes it possible to enlarge numbers, letters, words, symbols, and signs. When one of the lamps at the rear of the readout is lighted, it illuminates the corresponding film message, focuses it through a lens system, and projects it onto the nonglare viewing screen at the front. This "one lamp per message" design eliminates character misreadings caused by partial failures. Because the message displayed is on a single plane, there is no obstruction or confusion caused by unlighted filaments.

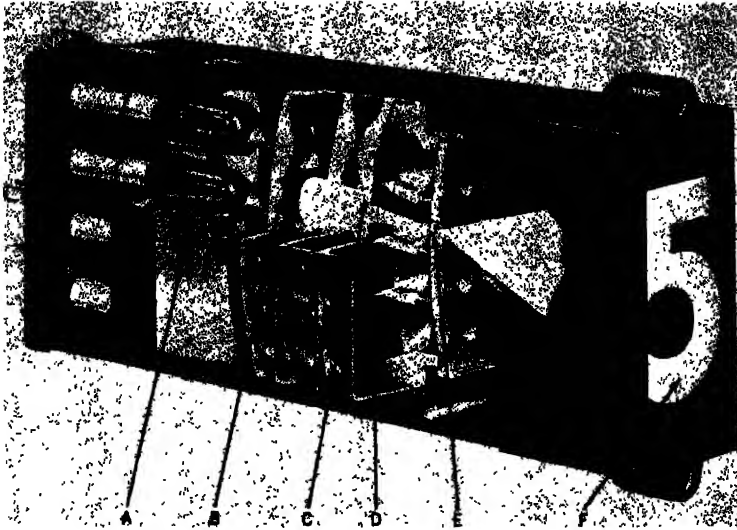


Figure 13. Projection Readout tube. (a) Standard MS or commercial lamp; (b) light-collecting lens; (c) dual square lens condensers provide greater coverage at lower magnification; (d) film containing display symbol (numbers, letters, words, symbols, colors); (e) projection lens; (f) nonglare viewing screen. (Courtesy of Industrial Electronic Engineers, Inc.)

3.5. AVAILABLE HARDWARE

So far we have discussed the three basic types of encoders, some of their virtues and limitations, and a few of their design problems. This section examines commercially available hardware. The risk involved in describing available instruments is that they will be obsolete before the ink is dry on this manuscript. Nevertheless, the principles described here illustrate the methods used in developing modern encoders. Not all encoders or manufacturers are covered. Only particularly interesting units were selected. The encoders are presented in the same order as that used previously. This in no way indicates relative merit.

3.5.1. Brush Encoders

As already stated, brush encoders are considered to be in the low and medium accuracy range but are relatively inexpensive. To compete with the more exotic wares, one manufacturer decided to make them more

attractive by combining the encoder with appropriate solid-state decoding networks and visual readout devices in one completely engineered package. This eliminates about 90% of the engineering headaches associated with encoder systems. Most commercial systems' requirements are compatible with this technique. The system is called Decitrak* and consists of three components: the Transmitter (encoder), the Gate Array (solid-state decoder), and the Read-Out or Print-Out. The Transmitter accepts mechanical input information via a gear or other coupling. Angular displacement of the transmitter shaft results in a direct decimal, electrical output. Two stationary disks contain a total of four patterns, one for each digit in the output. A brush rides on each pattern. Two brushes rotate on the coarse disk and generate the highest-order and next-highest-order digits; two brushes rotate on the fine disk to generate lower-order digits. The brush block holding the lower-order brushes rotates with the input shaft. The other brushes rest in a block that is geared to the input so that, at the end of every revolution of the input shaft, they advance $1/N$ revolutions (N = number of revolutions).

The Gate Array, containing a card file of solid-state logic circuitry, eliminates ambiguity, decodes, and provides the driving power for the readout device. The estimated MTBF is 15,000 hours.

The Read-Out is some form of nixie tube or a printer. Printers with 4-digit output, as well as units with a time base, are available (Figure 14). Standard units read to 4 or 6 digits. Resolution therefore is one part in 10,000 to one part in one million. Maximum continuous shaft speed is about 500 rpm. Breakaway torque varies from 0.2 to 20 in.-oz; life is up to 25×10^6 revolutions.

3.5.2. Magnetic Encoders

A typical, good magnetic encoder is made by Electro-Mechanical Research, Inc. It contains a ferrite disk, fabricated from a magnetic ceramic material of high coercive force, magnetized so that a flux pattern representing the Gray code is set up around 8 concentric tracks. The pickup heads—one for each track—are positioned close to the surface of the disk. Each rectangular loop toroid pickup has an input or interrogation winding plus an output signal winding. The operation of this encoder is described in Section 3.1.3. Normally the interrogation windings of the pickup heads are connected in series, and the output windings in parallel. A transistorized oscillator can be used to energize the interrogation winding. The waveform can be a sine wave, square wave, or even a pulsed wave so long as the input signal has alternative positive and negative excursions of sufficient amplitude

* Theta Instrument Corporation, Fairfield, N.J.

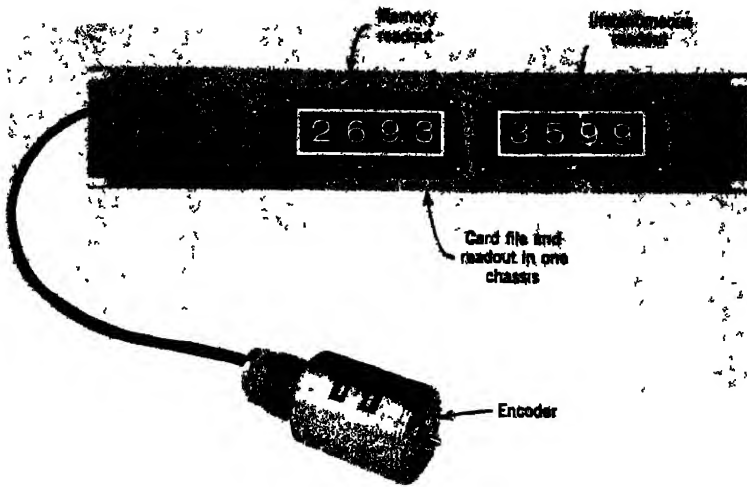


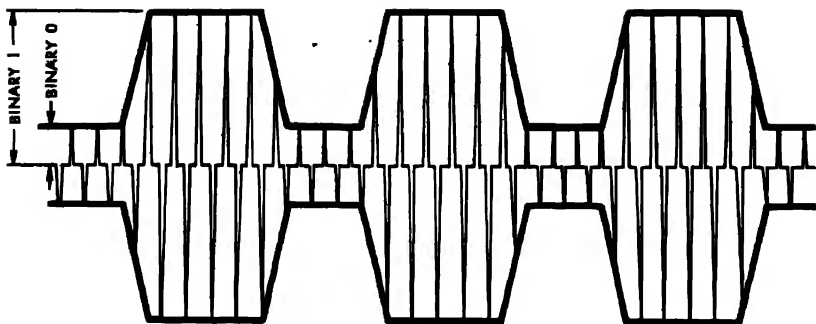
Figure 14. Brush encoder system. (Courtesy of Theta Instrument Corporation.)

to switch the toroid. The frequency of the signal appears in the output-pulse envelope as the carrier wave. Frequencies of 40 to 50 kHz are considered optimum. Figure 15a illustrates a typical output waveform. Figures 15b and 15c show an input signal and a corresponding output signal when the pickup head is not in a magnetic field. Figure 15d shows the corresponding Gray code pattern. Resolution is one part in 2^8 . Life is about 6×10^9 revolutions at 10,000 rpm. Breakaway torque is 0.05 in.-oz.

An interesting variation on this theme is the hybrid Magnetic-Brush encoder produced by the Librascope Group of Singer-General Precision, Inc. It utilizes a high-speed magnetic encoder stage coupled to one or more brush encoder stages through step-down gearing between stages. Life is increased over that of a brush encoder by a factor of 64, while only a portion of the electronics normally needed for a magnetic encoder is required. Brush encoders normally rated at a 2 million revolution life at 200 rpm can therefore be used to achieve a 128 million revolution life at 4000 rpm. Breakaway torque is about 0.3 in.-oz.

3.5.3. Optical Encoders

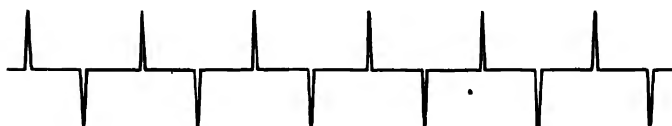
Optical encoders are considered to be the top of the line in today's encoder market. Like all good instruments, they have evolved from crude



(a)



(b)



(c)



(d)

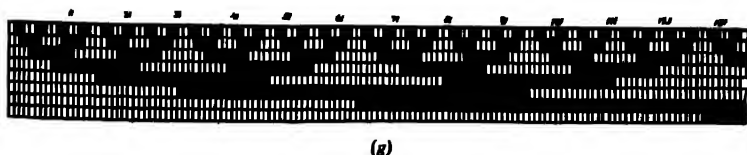
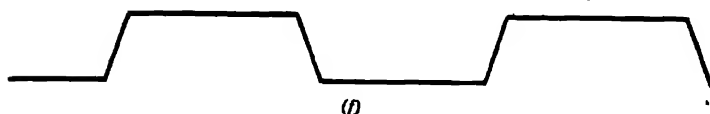
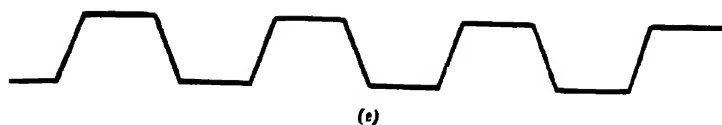


Figure 15. Magnetic encoder system. (a) Typical output waveform envelope modulating a 40- to 200-kHz carrier. (b) Input signal. (c) Output signal of head not in a magnetic field. (d) Output from LSB brush. (e) Output from next LSB brush. (f) Output from MSB brush. (g) Gray code pattern. Shaded areas represent binary "1." Unshaded areas represent binary "0." High signal output of unsaturated reading heads represents binary "1." Low signal output of saturated reading heads represents binary "0." For 2^7 operation use tracks 2 through 8 only; for 2^6 operation use tracks 3 through 8 only; and so on. (Courtesy of Electro-Mechanical Research Corporation.)

designs to exotic space-age concepts. You can buy encoders in various stages of development. The more exotic instruments are also, of course, the more expensive and the least tested in industrial usage.

The Baldwin Optical* encoder is a typical, good, incremental encoder that has evolved from a design once considered quite exotic. This instrument uses a conventional subminiature bulb with a photocell detector. It is available with outputs up to 48,000 counts per revolution. Typical output characteristics are shown in Figure 16. One of the most interesting aspects

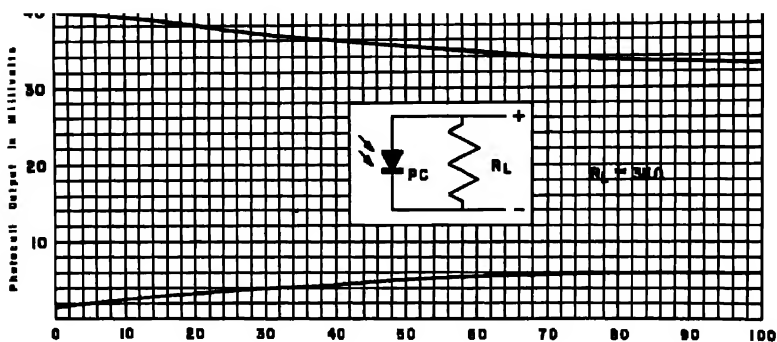


Figure 16. Optical encoder photocell output versus frequency. (Courtesy of Baldwin Electronics.)

of this class of encoder is the variety of electronic options available as an integral part of the encoder. This choice makes possible many unique systems concepts. One option is to utilize a Schmitt trigger for a very clean square-wave output free from hash and overshoots. Another is a pulsed output at positive slopes of the output waveform with a very precise pulse duration. It is also possible to multiply the number of pulses per revolution electronically so that the error of ± 1 count is further reduced. This can be combined with an integral counter that provides a zero reference pulse after a prescribed number of pulses have been tallied. Typical starting torque is 3 in.-oz with a life of 50×10^9 revolutions at speeds up to 30,000 rpm. One of its most common system applications is as a very precise tachometer.

An example of an exotic optical encoder is the DIGISEC made by the Wayne-George Company. It was designed to provide 19-bit accuracy—approximately one part in a half-million. The code disk in this encoder is

* Baldwin Electronics, Little Rock, Ark.

about 4 in. in diameter and has 15 concentric tracks, each composed of alternate opaque and transparent segments. The outermost track contains 2^{14} parts and is called the fine track. Each of these elements is about 0.0004 in. wide. Immediately above the code disk is an incandescent light source, while a slit-detector system is located below the disk.

By a suitable combination of code disk element shape, readout slit configuration, and photosensing geometry, the output signal of the photo-sensitive detector produces an output whose amplitude is nearly a sine wave function of input shaft angle displacement with a period corresponding to the width of one pair of opaque and transparent code disk elements. One complete revolution of the input shaft produces 2^{14} cycles of a sine wave.

$$\text{output} = K \sin [2\pi 2^{14}\theta]$$

where K = proportionality constant

θ = shaft angle

With a second slit array displaced $(n + \frac{1}{2})$ code cycles away from the first, a second output proportional to $\cos [2\pi 2^{14}\theta]$ is generated. Each photodetector signal is amplified and fed to a DC flip-flop that changes to its "0" state for negative inputs and to its "1" state for positive inputs (Figure 17). The sine function signal determines the state of the "A" flip-flop, and the cosine function signal determines the state of the "B" flip-flop. So far one revolution of the input shaft has been divided into 2^{14} parts. The next step is to divide the cycle into four parts. The four quadrants of a cycle can be identified by a 2-digit natural binary sequence as follows:

Quadrant	X_1	X_0
1	0	0
2	0	1
3	1	0
4	1	1

where X_1 = most significant bit

X_0 = least significant bit

Note that X_1 is a "0" in the first two quadrants and a "1" in the third and fourth. This is the pattern of the "A" flip-flop in Figure 17. We can therefore state that

$$X_1 = A$$

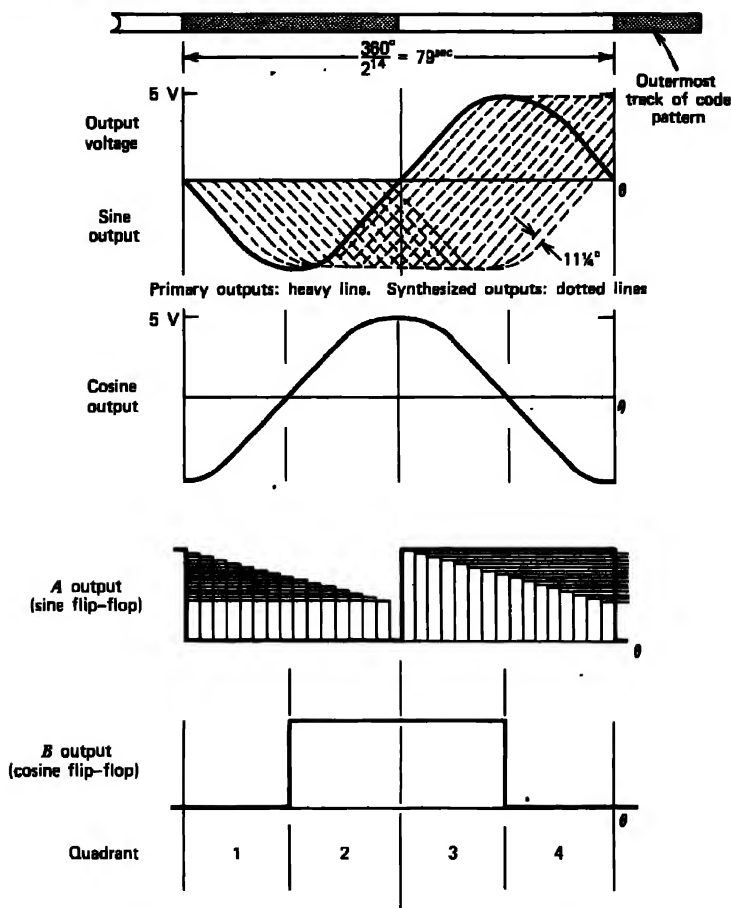


Figure 17. Digisec encoder—principle of operation. (Courtesy of Wayne-George Division, Ittek Corporation.)

and the “A” flip-flop output becomes the bit X_1 . Similarly, X_0 is a “1” in the second and fourth quadrants where A and B have opposite states. The X_0 can be represented by an “exclusive OR” as follows:

$$X_0 = \bar{A}B + A\bar{B}$$

By this logic, each of the 2¹⁴ cycles is divided into four equal parts for a total resolution of 2¹⁶. It is still necessary to extend this technique so that

the total subdivision is 2^5 instead of only 2^2 . This is accomplished by synthesizing additional sine and cosine functions of the angle by combinations of the primary sine and cosine waves. For example, analog addition of the sine and cosine waves with equal summing coefficients results in a new wave out of phase by 45° with respect to the primary sine wave. In a similar manner, sine waves of intermediate phase displacements from the primary sine wave can be synthesized by adding the primary sine and cosine waves with unequal summing coefficients. The net result of this process is the generation of a family of sine waves each displaced from the primary sine wave by a multiple of $360^\circ/2^5$ (Figure 17). So far, one rotation of the shaft is divided into 2^{19} parts, each equal to 2.47 arc-sec, by means of a single code track called the fine track. The 5-bit binary word becomes the 5 least significant bits of the 19-bit output. It is still necessary to identify each group of the 2^{14} cycles of the fine track. This is done by using 14 additional tracks, laid out in a conventional binary pattern, each of which provides 1 bit. A carry logic technique, similar to V-Scan, determines the selection of a lead or lag detector on each track depending on the state of the next-lower-order bit.

The DIGISEC achieves high accuracy by two readout techniques. First, the outer track is read through a large number of slits; as many as 300 adjacent code elements are read simultaneously. This procedure not only increases signal output, but also provides substantial averaging which significantly reduces local code element errors. Second, four readout stations spaced at 90° intervals are used. This not only increases averaging by a factor of 4, but also reduces the effect of disk pattern error. For example, two stations located 180° apart eliminate the effects of bearing run-out and code pattern eccentricity. Four stations also eliminate any second harmonic component of error due to ellipticity in the code pattern.

Typical mechanical characteristics include breakaway torque of 0.15 in.-oz and a life of 10^9 revolutions at 5000 rpm. See Figure 18 for a handy conversion chart of bits and their decimal equivalents.

Another unique encoder is the Optisyn made by the Dynamics Research Corporation. Harold H. Seward of the MIT Instrumentation Laboratory originally developed the Optisyn principle for use in measuring displacements in inertial navigation systems. It is best described as an optical gearing technique; a very small angular displacement of the input shaft produces a large displacement of the encoder pattern. The Optisyn uses two transparent disks having photographic patterns of alternately opaque and clear sectors. One disk has one more opaque and transparent sector than the other disk, resulting in a moiré fringe pattern (Figure 19). The two disks are mounted concentric to the axis of rotation, with one disk fixed to the case and the other mounted on the input shaft. The standard Optisyn has four lamps

DECIMAL	BINARY	ANGLE	
1	2 ⁰	360	DEG.
2	2 ¹	180	
4	2 ²	90	
8	2 ³	45	
16	2 ⁴	22.5	
32	2 ⁵	11.25	
64	2 ⁶	5.625	
128	2 ⁷	2.8125	
256	2 ⁸	1.40625	
512	2 ⁹	42.1875	MIN.
1,024	2 ¹⁰	21.0938	
2,048	2 ¹¹	10.5469	
4,096	2 ¹²	5.27344	
8,192	2 ¹³	2.63672	
16,384	2 ¹⁴	1.31836	
32,768	2 ¹⁵	39.5508	SEC.
65,536	2 ¹⁶	19.7754	
131,072	2 ¹⁷	9.88770	
262,144	2 ¹⁸	4.94385	
524,288	2 ¹⁹	2.47192	
1,048,576	2 ²⁰	1.23596	
2,097,152	2 ²¹	0.617981	
4,194,304	2 ²²	0.308990	
8,388,608	2 ²³	0.154495	
16,777,216	2 ²⁴	0.077248	

Figure 18. Decimal-to-binary conversion chart. (Courtesy of Baldwin Electronics.)

mounted at 90° intervals on one side of the disks; four corresponding phototransistors are mounted in alignment on the other side. The lamps transmit light through the interfering disk patterns to the phototransistors. For a given relative orientation of the disks, a minimum amount of light will be transmitted in one region around the disk circumference. At 90°, on either side of this region, the clear sectors will be half open, and at 180° the clear sectors will be aligned and completely open for maximum transmission of light. If the Optisyn shaft is rotated by an amount equal to one sector angle, the dark and light transmission regions rotate 180°. A full rotation of the movable disk causes the dark and light patterns to rotate n times; n is the number of opaque lines on the rotating disk. The optical light pattern is thus "optically geared" to the input shaft with a magnification ratio of n .

If a DC voltage is applied to the instrument, the output waveform will be triangular. One pair of illuminated phototransistors produces a single waveform with positive and negative phases. Two pairs of phototransistors,

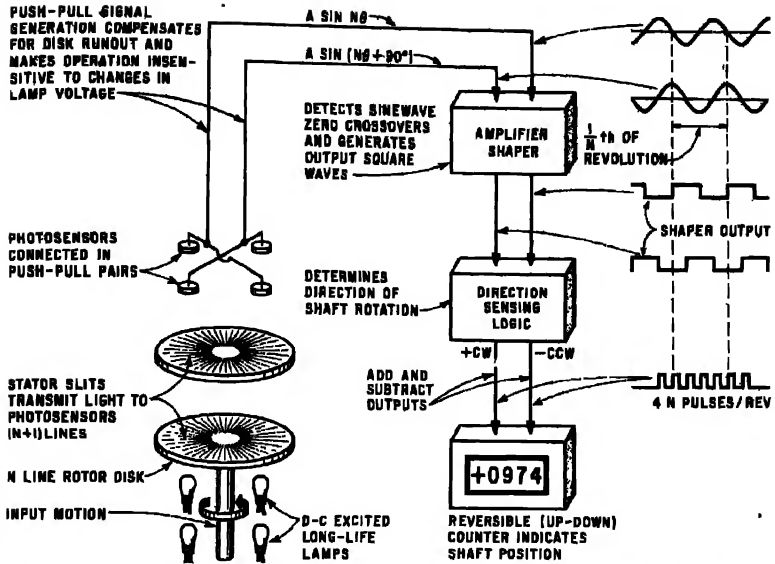


Figure 19. Optisyn encoder—operating principles. (Courtesy of Dynamics Research Corporation.)

mounted at 90° intervals, produce two triangular waveforms 90° out of phase. If an AC supply is used, the same waveform will be produced in the form of the envelope of modulation. The AC output signal is similar to that of an ordinary synchro except that the modulated envelope is triangular rather than sinusoidal and occurs n times per revolution rather than only once. For two diametrically mounted and interconnected transistors, the net output voltage goes positive or negative, depending on whether one or the other transistor receives more light from the rotating pattern. In one Optisyn, the linear waveform could be read to within $\pm 5^\circ$. This accuracy combined with an optical gearing ratio of 512-1, results in a shaft angle accuracy of better than $1/30,000$ of a revolution. Optisyns with counts up to 2^{14} per revolution are available. Mechanical performance include starting torques down to 1 in.-oz, a life of 2×10^8 revolutions with maximum speed of 5000 rpm.

The Dynamics Research Theodosyn encoder is specifically designed for applications where neither conventional encoders nor Optisyns can provide the required resolution and accuracy. The Theodosyn is an incremental optical shaft encoder containing an optical system that superimposes the pattern from one side of its code disk onto the pattern on the opposite side

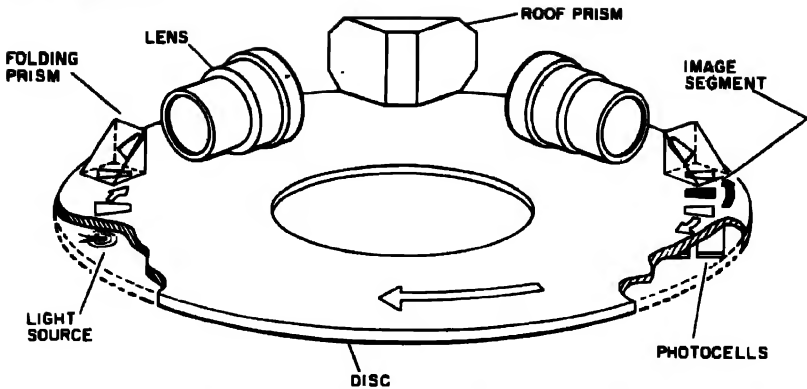


Figure 20. Theodosyn encoder—operating principles. (Courtesy of Dynamics Research Corporation.)

(Figure 20). The basic instrument consists of a glass disk divided into equally spaced clear and opaque sectors. The light source is under one side of the disk; photocells are diametrically opposite. The optical system—two lenses, two 45° prisms, and a roof prism—projects the illuminated disk segment to the photocell side of the disk. The code track consists of 2^{15} alternately clear and opaque radial sectors. Optical superimposition of segments produces an interference pattern which alternates between maximum and minimum light intensity at twice the frequency of the lines on the disk. From this interference pattern, each of two quadrature-phased photocell channels generates a 2^{16} cycle per revolution semisinusoidal output waveform. Discrimination of the two null crossings per cycle of the signal from each of the two channels provides angular resolution of 18 bits per shaft revolution. Additional phase-shift circuitry can be provided to extend the resolution to 20 bits. This design virtually eliminates the errors that ordinarily result from shaft misalignments and disk ruling imperfections. Since the signal is generated by the optical superimposition of diametrically opposite sectors of the rotating disk, it does not need stationary reticles used in other encoders. Starting torque is about 0.300 in.-oz. Life is 2×10^8 revolutions at an operating speed of 60 rpm.

3.5.4. Sine-Cosine Encoders

Many applications require the computation of range or distance in solving a problem. It is feasible to monitor the angular shaft deflection of a conventional encoder and then convert the readings to sine and cosine functions, but a direct reading in these functions is certainly more con-



Figure 21. Sine-cosine encoder disk. The sine-cosine encoder is a multitrack absolute device. The total output is the sum of the voltages across each brush. The disk pattern is arranged so that the waveform generated is essentially a sine wave with a series of very small discrete steps. The cosine waveform is generated by a group of brushes spaced 90° away from the brushes generating the sine wave pattern. (Courtesy of W. & L. E. Gurley Company.)

venient. When we obtain the function directly, it eliminates the requirement for a computer (Figure 21). Present optical models are available with 14-bit accuracies. Other characteristics are similar to the basic units described above.

3.5.5. Linear Encoders

Linear encoders are used for the conversion of rectilinear motion into incremental, direction-sensed, electrical pulse outputs. The basic methods

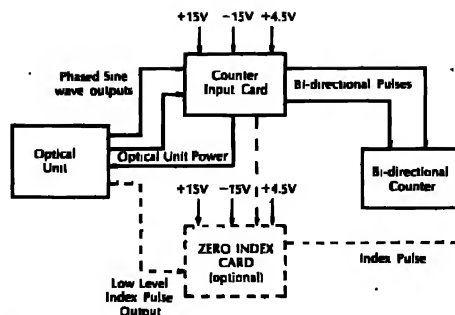


Figure 22a. Linear encoder schematic. (Courtesy of Wayne-George Division, Itek Corporation.)

of generating these pulses is similar to those employed in shaft encoders. The encoder is mounted on a fixed surface and the coupling member is attached to the moving part, such as a machine tool carriage or an *X-Y* plotter. The instrument will generate a single-pulse output for each increment of displacement depending on the direction of motion of the moving member. To keep track of the net encoder displacement from the starting point, the pulses are tallied in a bidirectional counter. Since the encoder is an incremental device, starting points are established simply by resetting the counter to zero. By using a counter with a visual decimal readout, the system can easily be read to 50 μin . One such unit, made by the Wayne-George Corporation, consists of a precision glass code plate and an optical reading head containing the lamps and photodetectors (Figure 22a). Motion of the reading head relative to the glass code plate generates two 90°-phased sine waves whose length is determined by the width of alternate transparent and opaque segments on the code plates. The Counter Input Card combines the two 90°-phased sine waves in controlled ratios to produce a family of precisely-phased sine waves, which are first squared and then differentiated. The resulting pulses are then combined with the square waves from which they were derived in direction-sensing logic circuits to produce the two bidirectional output pulse trains. Typical units can monitor motion up to 10 in. at a maximum velocity of 10 in./sec. Life exceeds 30,000 hours of operation. Another version of a linear encoder is shown in Figure 22b.

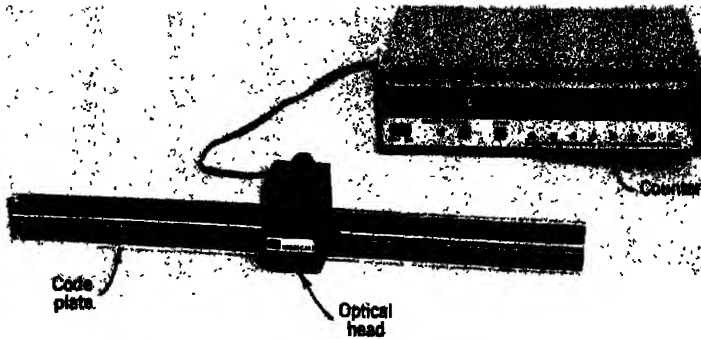


Figure 22b. Linear encoder. (Courtesy of Dynamics Research Corporation.)

3.6. ENCODER APPLICATIONS

The encoder is a basic sensor in digital servosystems. It has two modes of operation:

1. As a position indicator.
2. As a velocity indicator.

Each mode of operation can be further subdivided into open and closed loop applications.

When an encoder is used as a position indicator it is comparable to a synchro. The important difference is that the synchro is an analog device with an output proportional to the sine of the displacement angle. Since the synchro is an analog device, its output voltage is immediately useful to conventional servocomponents such as amplifiers, control transformers, and control differentials. An encoder, however, must be used in conjunction with a digital-to-analog converter to be useful with conventional servo equipment. Although this arrangement is more expensive than conventional servos, it is more accurate (Figure 23). When an open-loop system is used, no such complications exist; the output of the encoder is simply registered or displayed on digital devices.

When a simple indication of shaft velocity is required, the output of an incremental encoder is fed to an event-per-unit-time device. For closed-loop applications, where the encoder is replacing a tachometer, the output of the

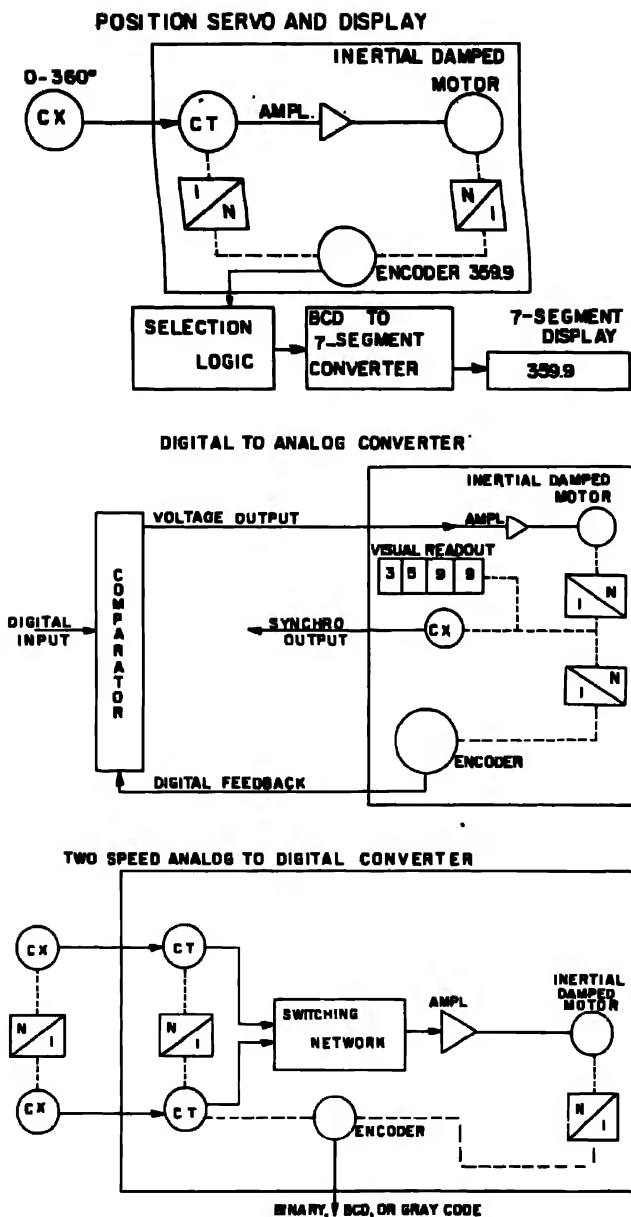


Figure 23. Encoder applications in analog and digital systems. (Courtesy of Singer-General Precision, Inc.)

encoder is wired in shunt with an appropriate load. The resultant output can be used for closed-loop control functions. The following examples are typical of current practice.

3.6.1. Brush Encoder Applications

The DECITRAK system is a good example of open-loop brush encoder systems. The encoder can be coupled or geared to any type of sensor where the output is in the form of shaft rotation (Figure 24). The encoder (transmitter) output is then converted to digital form and can be used to

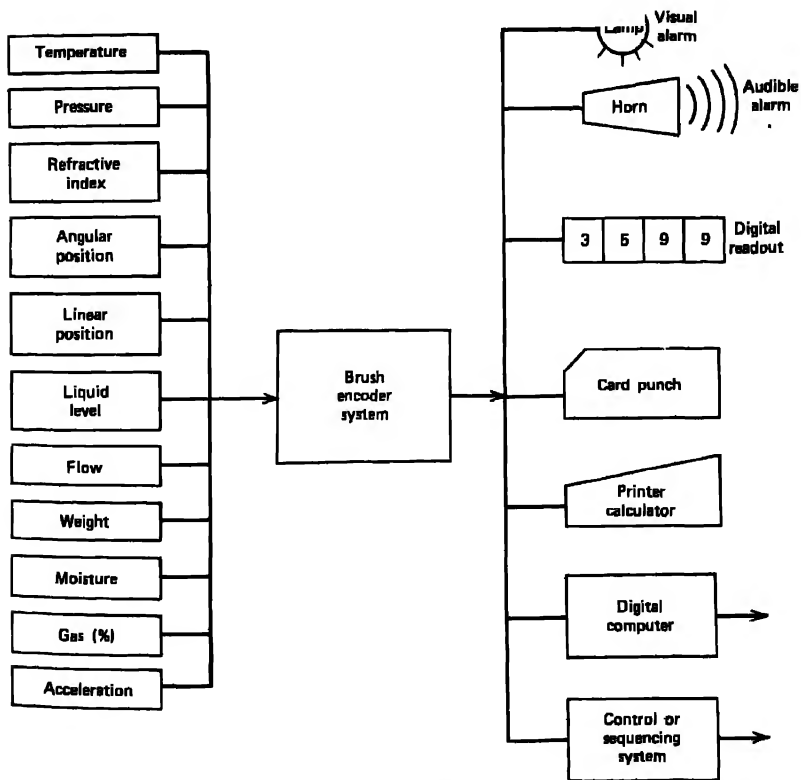


Figure 24. Decitrak brush encoder system. (Decitrak is a trademark of the Theta Instrument Corporation.) (Courtesy of Theta Instrument Corporation.)

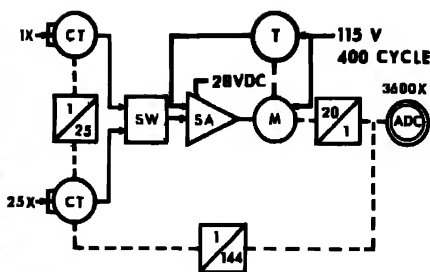
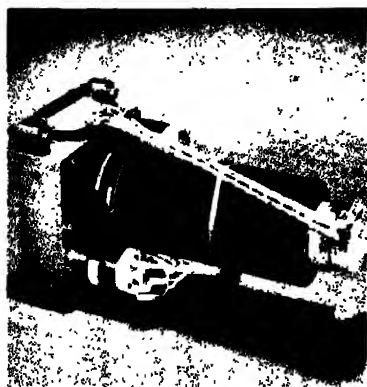


Figure 25. Magnetic encoder used in a navigational servo for monitoring latitude. CT = Control transformer; SW = selector switch; SA = servoamplifier; M = servomotor; T = tachometer; ADC = magnetic encoder. (Courtesy of Norden Corporation.)

drive numerous readout devices. Brush encoder systems are 10 to 1000 times more accurate than comparably priced analog systems. Compared to optical encoders, they are less accurate but also less expensive. They are not capable of the high shaft speeds of magnetic encoders, but they require less circuitry and are also impervious to the effects of noise and pickup.

3.6.2. Magnetic Encoder Systems

Magnetic encoder systems differ from brush encoder applications chiefly in their capability of monitoring much faster shaft speeds and in their exceptionally long life. They are also generally available with antiambiguity devices not found in many brush encoders. Bidirectional readout data are also available. The greatest use of magnetic encoder systems has been in the aircraft instrument field. Although the encoder is normally used as a primary sensor or as a tachometer, it is also useful for extending the resolution of synchros (Figure 25). The system shown is made by the Norden Corporation for measuring latitude or longitude in aircraft. The primary sensor is a control transformer. The output of the motor shaft, which is proportional to the synchro (CT) output, is directed through a gear train to a magnetic encoder. In this way the position of the motor shaft and synchro shaft can be monitored more accurately than with conventional techniques. The module illustrated is equipped with a tachometer for servo stability and

with an option switch that provides either direct coupling or a 1-25 gear ratio between the synchro and the airplane component being monitored. The encoder used generates the latitude function, complete with sign, in degrees and hundredths of degrees. Similar units are made for monitoring altitude, heading, and ground speed.

3.6.3. Optical Encoder Systems

State-of-the-art applications of optical encoders provide ultraprecise position and velocity control of industrial processes. To illustrate the approach, one system is discussed in detail here. This system was designed to produce control accuracies of 1 ppm even when experiencing variations in supply voltage, load torque, and environmental conditions.

The Sequential FPL* system uses an optical encoder as a tachometer, whose output is compared against a reference frequency to produce velocity and position control (Figure 26). A reference frequency is directed through

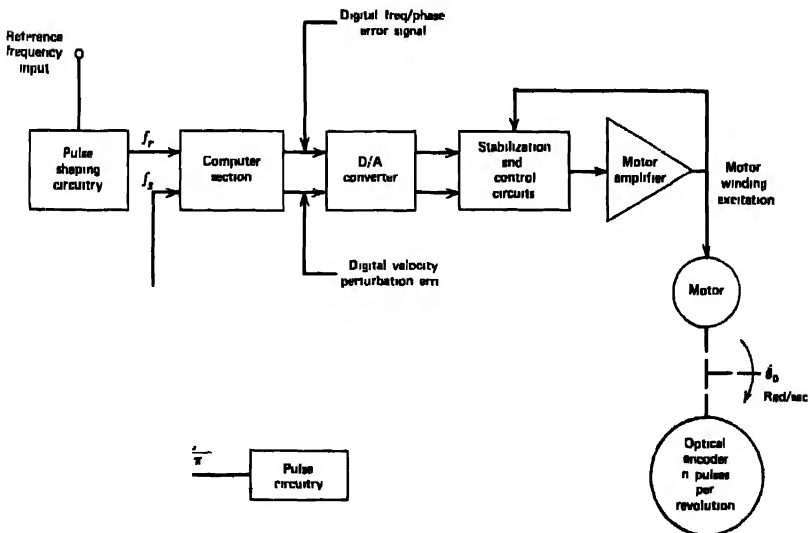


Figure 26. Optical encoder application in a control system. (Courtesy of Sequential Electronics Corporation.)

* Trademark of the Sequential Electronics Systems, Inc., Dobbs Ferry, N.Y.

pulse-shaping circuitry to a comparison computer. The computer compares the reference signal f_r to the signal produced by the encoder f_e . The resultant error signal is converted from a digital to an analog voltage and is used to drive the motor. The encoder monitors the motor speed and supplies a feedback signal through pulse-shaping networks to the computer. To maintain very precise control, the computer not only measures the difference in frequency between the two signals, but also compares the relative phase angles. When frequency lock has been achieved, a digital error signal proportional to phase angle is generated. The frequency and phase lock circuits operate simultaneously, and the transition from one mode to the other is performed automatically in a smooth manner so that no switching transients appear in the error signal. The D/A converter converts the difference of the two signals impressed to a DC voltage on a pulse-to-pulse basis. The electrical bandwidth achieved is one half the frequency of the reference signal; hence the more pulses produced by the encoder per revolution, the higher the reference voltage frequency and the better the frequency response.

The control section contains stabilization circuitry that applies phase, velocity, and acceleration control in the proper ratio, and with proper frequency characteristics, to optimize the closed-loop performance of the system. The amplifier unit supplies the correct excitation to the motor windings to maintain constant shaft velocity. Excitation signals from the amplifier are fed back to the control section, so that the amplifier and motor windings are operated in a secondary, closed-loop manner. This effectively linearizes the torque-output characteristic of the motor and almost eliminates the effect of the electrical time constant of the motor winding. If a torque disturbance is applied to the motor shaft, the phase angle between the encoder and reference signals will shift, changing the phase error signal. The torque output of the motor will change just enough to cancel the disturbance and maintain speed synchronization. The maximum electrical phase shift is limited to $360/K_p$ degrees, where K_p is the electrical phase-lock gain. Since the encoder produces N pulses per revolution, $360/K_p$ electrical degrees corresponds to $360/NK_p$ mechanical degrees. The mechanical phase-locking accuracy is therefore a function of encoder resolution and electrical phase-lock gain. Typical values of K_p are 20 to 40 dB with encoder resolution up to 2^{19} parts. This produces closed-loop positional accuracies in the order of seconds of arc. Typical velocity errors are better than 0.001% from 0 to 1 kHz and about 0.02% from 1 to 20 kHz. The upper frequency response of a typical system is about 1 MHz. Time displacement errors are 100 nsec at worst.

This phase-lock technique is used extensively in many very accurate servosystems, such as film synchronization systems, magnetic storage drums, and incremental position controls.

REFERENCES

1. Shaft Encoders, *Electromechanical Design*, January 1967, Benwill Publishing Company, Brookline, Mass.
2. D. E. Hoeschele, Jr., *Analog-to-Digital/Digital-to-Analog Conversion Techniques*, John Wiley & Sons, Inc., New York, 1968.
3. *Singer-General Precision, Inc./Librascope Encoder Data Sheets*, Glendale, Calif., 1968.

Chapter IV Timers

The development of timers has paralleled the advance of science throughout history. The sundial came into being during the period when studies of geometry were initiated. Hourglass timers were typical of the Middle Ages. The first clock developed by Pope Silvester II in 996 was indicative of the first stage of Renaissance thinking. During the next 600 years improvements came slowly as new fabrication techniques in wood and then in metalworking evolved and metallurgy became a science. The first timers, used to control rather than to measure time, appeared at the beginning of the twentieth century as electricity became an industrial tool. Pneumatic controllers came into use soon after the development of steam engines and compressors, but the modern versions date back to the World War I era. Electronic timers are only about 20 years old; the general utilization of solid-state circuits and time measured in nanoseconds is less than 10 years old.

Timers are grouped in six categories:

1. Clocks and escapement devices.
2. Pneumatic instruments and dashpots.
3. Electrothermal instruments.
4. Electromechanical units and relays.
5. Magnetically controlled devices.
6. Electronic instruments.

4.1. CLOCKS

Clocks no longer have system applications; escapements, however, are used as integrators in systems in which no power is available or in situations where an instrument must continue to function during a power failure. They are discussed in Section 4.7.

4.2. PNEUMATIC DEVICES

Pneumatic timers are based on the time delay obtained by a dashpot. A dashpot is basically a tight-fitting piston and an enclosing cylinder where air, or another fluid, must pass between the radial clearance around the piston before the piston can move. Unlike an automobile cylinder, a dashpot is not intended to compress air but rather to meter its flow (Figure 1). The

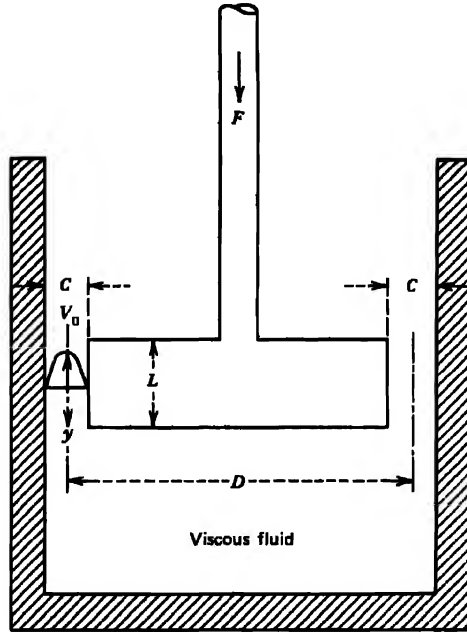


Figure 1. Dashpot. (Courtesy of McGraw-Hill Book Company. From Reference 11.)

movable part of the dashpot is generally attached to a set of contacts so that when the unit completes its stroke a circuit is completed or opened. Although the primary objective of a dashpot in a pneumatic timer is to produce a time delay, the design approach starts with its damping factor. The damping factor, f , may be expressed as follows:

$$f = \frac{F}{V} = \frac{3}{4} \mu \pi L \left(\frac{D}{C} \right)^3 \quad (1)$$

where f = damping factor (pound-seconds per inch)

F = damping force (pounds)

V = average piston velocity (inches per second)

μ = absolute viscosity of the fluid (pound-seconds per square inch)

L = length of piston (inches)

D = effective diameter of the piston (inches) (see Figure 1)

C = radial clearance (inches)

This equation shows that the damping factor is directly proportional to the viscosity of the fluid used in the dashpot. The viscosity is in turn a function of temperature; consequently, dashpots are highly dependent on ambient temperature variations. Good design practice normally limits the effects caused by variations in the length and radial clearance of the piston. Practical production tolerances make it necessary to provide some means of adjusting the damping factor. A standard solution is to provide a bypass equipped with a variable orifice (Figure 2). The orifice has a fine-adjusting

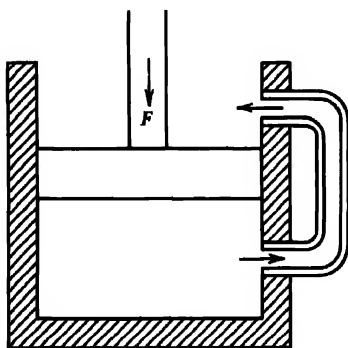


Figure 2. Dashpot with bypass orifice. (Courtesy of McGraw-Hill Book Company. From Reference 11.)

screw so that the degree of damping can be precisely regulated. One design uses a bimetallic element to open and close the bypass as the temperature varies. Some dashpots have the bypass in the end of the cylinder so that some of the fluid may escape through the end of the unit (Figure 3). So far the analysis deals with a dashpot in the vertical plane where the piston moves downward under the force of gravity. If the unit were tilted the time delay

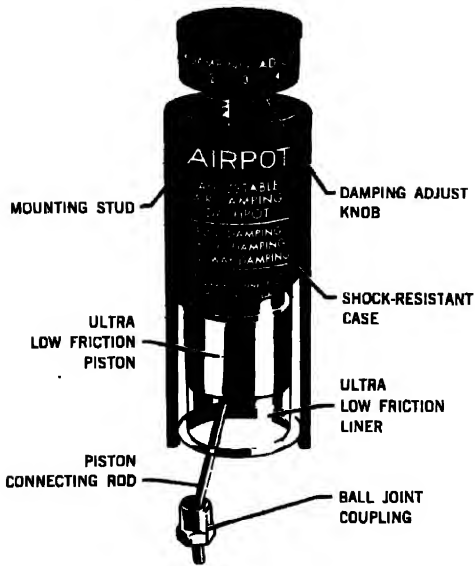


Figure 3. Commercial dashpot. (Courtesy of Electric Regulator Company.)

would change drastically. Commercial timers generally use a spring to load the piston of the dashpot so that the device is substantially free from the effects of the earth's gravitational field.

Although pneumatic timers are based on the dashpot theory described above, commercial instruments may digress widely from the basic theme. The unit shown in Figure 4 uses a flexible diaphragm to compress air. The flow of air is metered by a filter and nozzle combination. The cross section of the nozzle is adjustable to provide control of the time delay. The diaphragm is compressed by a solenoid-actuated plunger. Another variation of this type of design is shown in Figure 5. This instrument uses only a metering valve to regulate the time delay.

Pneumatic timers are medium-accuracy instruments; the time delay tolerance range is 5 to 10% but repeatability is much better. Time delays are available in ranges from a tenth of a second to several minutes. In case of power failure they always return to their initial position under spring pressure. This is not the case for other types of timers. A wide range of gases and fluids are available to vary the viscosity and the flexibility of these instruments. Many types of contact arrangements are available as catalog items. They are particularly useful when the current loads are high.

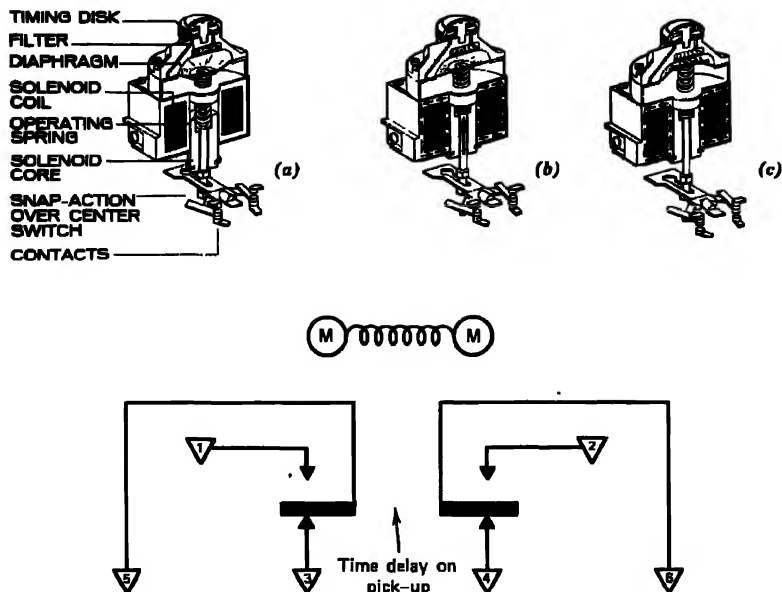


Figure 4. Pneumatic timer. This device provides delayed switch transfer on energization and instant reset on deenergization. (a) Normal deenergized condition: (b) Coil is energized, drawing solenoid core to upper position and compressing operating spring. Spring pressure on diaphragm forces air through filter and into circular groove of timing disk. Setting of calibrated dial controls effective length of groove, which determines time required to exhaust air from timing chamber, hence the length of the delay period. (c) At the end of the delay period snap-action switch is forced over center, "breaking" the NC contacts and "making" the NO contacts. The switch remains in this transferred position as long as coil is energized. When the coil is deenergized (at any time during or after the timing cycle), the one-way valve allows the unit to recycle instantly to its original condition. (Courtesy of Amerace-Enna Corporation.)

4.3. ELECTROTHERMAL DEVICES

Electrothermal timers are composed of a bimetallic element, a heater, and a set of contacts. In the simplest type of thermal timer, the heater is wrapped around the bimetallic element; as the element deforms under heating, a contact closure is made (Figure 6). Some of these devices have the other half of the contact pair on another bimetallic element, not equipped with a heater. This is to ensure that the gap between the contacts is a fixed distance as the ambient temperature varies. Other variations on this theme

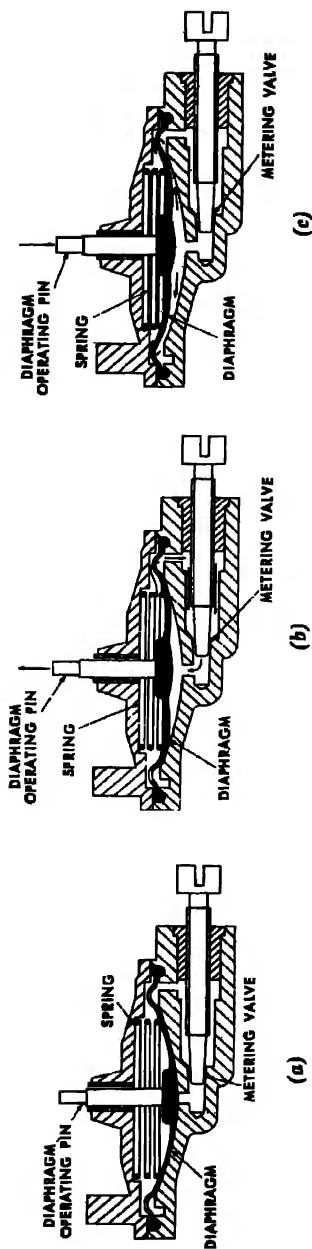


Figure 5. Pneumatic timer—metering valve type. (a) Reset position. The magnet, cavity, spring, and atmospheric pressure hold the diaphragm in complete contact with the sealing surface. Note that the operating pin is an integral part of the solenoid magnet assembly (not shown). (b) Timing stroke. The operating force draws the center of the diaphragm upward, displacing air from the spring cavity. Air circulates through perforations at the outer edge of the diaphragm through the metering orifice into the timing cavity. When sufficient air enters, the contacts transfer (contacts are not shown). (c) Risetting stroke. The force applied by the magnet defeats the atmospheric seal holding the outer portion of the diaphragm against the sealing surface. Air recirculates from the timing cavity through the perforations back into the spring cavity. On completion of the stroke, the diaphragm rest position is reestablished. (Courtesy of Cutler-Hammer, Inc.)

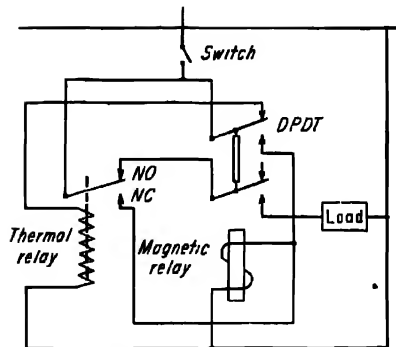
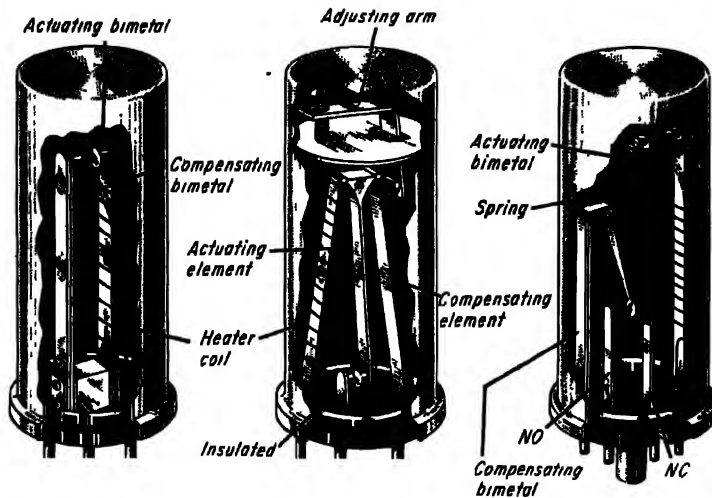


Figure 6. Thermal timers. (Courtesy of McGraw-Hill Book Company. G. V. Controls Co. From Reference 1.)

are also shown in Figure 6. As a system component this instrument has one principal flaw: it is prone to malfunction under vibration. Since the actuation elements are basically cantilever beams with relatively low natural frequencies, their performance under shock and vibration can be predicted. On the positive side, they have numerous uses in many noncritical industrial problems. They are low-priced and rugged and require very little maintenance. Some models have provision for adjusting the time delay over a limited range. Most of those instruments are hermetically sealed and filled

with a gas that will contribute to reliable performance. Some of the gases, such as hydrogen, neon, or argon, are designed to keep the contacts from oxidizing while providing reasonably good heat-transfer characteristics. One of the principal uses of thermal timers is as a delay when energizing equipment. To guard against starting transients being applied to sensitive circuits, the timer energizes the circuits when primary components have had time to stabilize. Since the time delay is a function of the power to the heater, which in turn is proportional to the square of the applied voltage, the timer provides a degree of protection against abnormal power supply conditions. One of the oldest applications for these devices is controlling flashing electric signs. Tolerance on the "on-off" time cycle usually runs about 10 to 20%.

4.4. ELECTROMECHANICAL DEVICES

Electromechanical timers are the most widely used timing instruments. The principle of operation is a motor driving a cam that periodically actuates a switch. The motor is generally a synchronous machine that is about as accurate as the stability of the exciting frequency; normally this is about 0.1%. The cams are available in a variety of shapes and materials. The most popular arrangement is two cams mounted face to face for each circuit. In this way one cam is used to trip the switch and the second cam to reset it. The most exacting part of calibrating a timer is the setting of this "on-off" period. If only one cam is used to perform this function, the accuracy of the timing cycle is limited to the machining tolerances on the lobe of the cam; when two cams are used, the machining tolerances are relatively unimportant. The adjustment of one cam relative to the other is the only important consideration (Figure 7). The ease of adjusting the timing period of an electromechanical timer is its most important advantage. The number of cams and circuits that can be handled is almost without limit. Some companies have 24 circuits in stock. The cams may be set so that each circuit has a fixed relationship to the other or one reference circuit (Figure 8). For systems work, a timing chart is designed before the timer is adjusted (Figure 9). The weak part of these timers has usually been the switches. They suffer from contact bounce, fatigue, and high contact resistance over prolonged usage. Initially, all these effects can be adjusted to optimum conditions but over a span of time that includes shock, vibration, and temperature cycling, the switches generally need repair before the motor or cams. Fortunately timers are no longer limited to conventional switches;

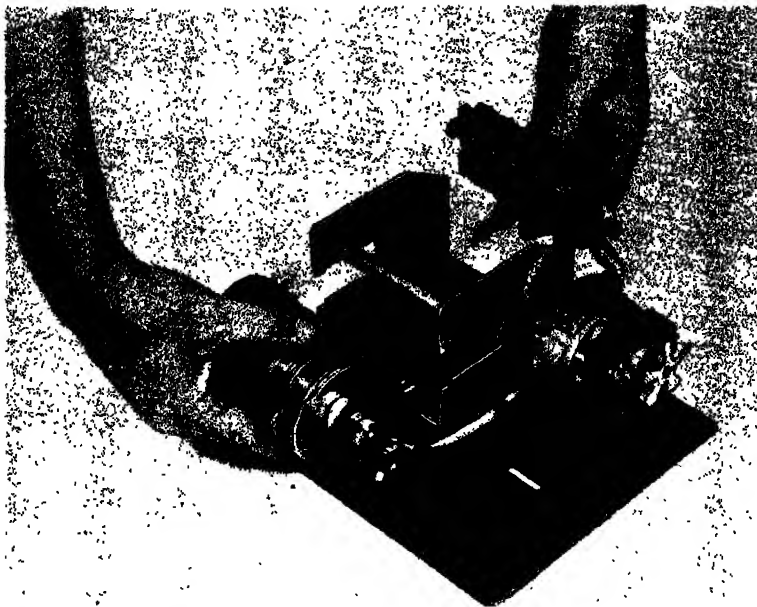


Figure 7. An adjustable electromechanical timer. (Courtesy of Bayside Timers, Inc.)

most of the noncontacting displacement devices discussed in Chapter 2 can be adapted for timer applications. The only problem is economics.

Current literature and trade magazines use five categories of timers that can be very confusing. The General Time Corporation has proposed the following definitions.

Repeat cycle timers are also called cam timers, program cam timers, or just cycle timers. They are the basic units described above and may be used whenever a sequence of events is to be regularly repeated.

Interval timers are used when the load is to be turned on and kept on for a specific interval and then turned off automatically. Typical applications are in ovens, dishwashers, and other automatic appliances. Most interval timers can be set externally to a wide range of values.

Elapsed time indicators show numerically the passage of time, similar to the way an odometer shows miles traveled. Basically they are integrators that may be called time totalizers or elapsed time meters. They are useful in recording the total running time of an electric device. The components are simply a synchronous motor driving a counter.



Figure 8. Adjustable cams make possible the adjustment of electromechanical timers. (Courtesy of Bayside Timers, Inc.)

Reset timers have a built-in mechanism which causes the timer to return to its starting position after it has measured a given time interval. They contain an integral motor-clutch unit. When the timer reaches the end of the interval, the clutch disengages itself from the output gears and the pointer on the dial is returned to the zero position under spring pressure.

Delay timers are used to provide the warm-up period in a copy machine, processor, or other electrical device. When the timer is activated, it measures the time interval, closes the power switch, and then deactivates itself.

Some manufacturers feature instruments that may be used for all five functions. If the application is only a simple time delay, a general-purpose timer may be an expensive solution. Typical hardware features time delays from a few seconds to 100 hours. Repeatability may be better than 0.1%. Interval accuracy is typically $\frac{1}{4}$ to 5%. When a timer is selected that uses an induction motor or DC motor, the accuracy and repeatability are not so good.

Another method of producing a time delay is the time-delay relay. It is discussed in Chapter 5.

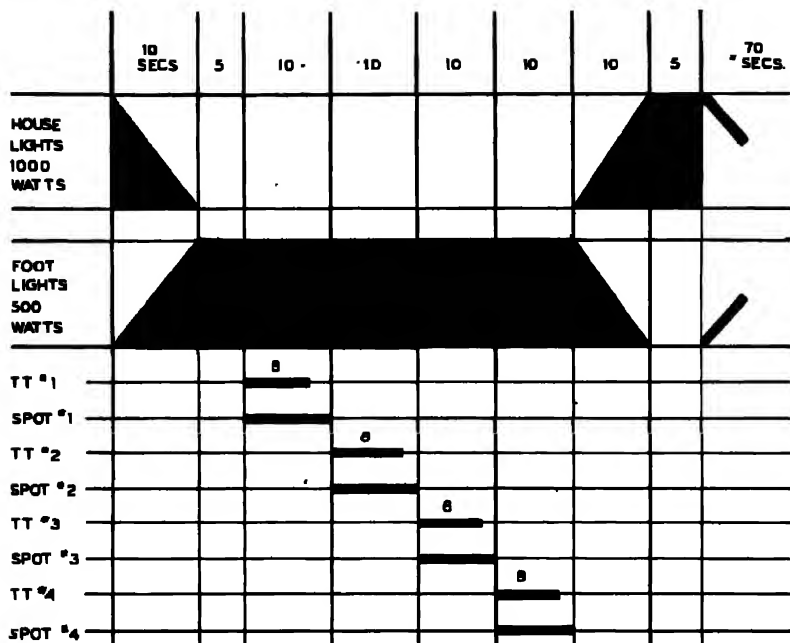


Figure 9. Timing chart. The timing chart of a hypothetical theater with four spotlighted turntables operated in chaser sequence. At the start, the houselights fade out as the footlights fade in. The change takes 10 sec. Five seconds later the first turntable is spotlighted as it starts to revolve, after 8 sec it stops, and 2 sec later its spot goes out as spot 2 comes on and turntable 2 starts. Each turntable in turn revolves for the first 8 of the 10 sec it is spotlighted. As turntable and spot 4 finish, the footlights go down as the houselights come up. The sequence is then repeated. (Courtesy of Bayside Timers, Inc.)

4.5. MAGNETIC DEVICES

Magnetic timers are composed of four principal components: an oscillator, magnetic core counter, logic, and an output element. The oscillator produces a series of pulses that are counted magnetically. When the proper number of counts has elapsed, the logic energizes the output device (Figure 10). The oscillator is a subminiature, temperature-compensated, low-frequency device. Low current drain enables it to operate from batteries when necessary.



Figure 10. Magnetic timer principle.

The chain of pulses is directed to a series of magnetic cores that count in much the same manner as a conventional electronic ring counter. When one series of cores is saturated, a pulse is directed to the next group. When all cores are saturated in a predetermined pattern, the logic circuitry energizes the output device. This may be a relay or some form of solid-state circuitry. The package size is extremely small; a timer with a 50,000-sec range may be packaged in only $1\frac{1}{2} \times 1\frac{1}{2} \times 3\frac{1}{2}$ in. volume. Overall accuracy is about $\pm 1\%$ with repeatability of $\pm 0.25\%$. Power consumption is less than 1 W.

4.6. ELECTRONIC DEVICES

Electronic timers are now almost as popular as electromechanical units. They are smaller, consume less power, and resist environmental effects better. Unfortunately, they are still more expensive and lack the flexibility of the motor-cam instruments. The principal virtue of electronic units is that they require no maintenance.

Note that there is no one circuit configuration in universal use today. Characteristically, the only limit on new circuitry is the imagination of the designer. Each design in turn has many variations to meet specific requirements. The circuits presented below are basic in nature and are available in many design manuals.

The heart of the circuit is a silicon-controlled rectifier (SCR). This is simply a bistable solid-state device. Its normal state is to block voltage; however, when the gate voltage is made positive with respect to the anode, the device becomes conductive (Figure 11). The SCR will remain in the conductive mode even if the voltage applied to the gate is applied for only 10 to 20 μsec . It requires only a "holding current" to remain in this state. This circuit is often called a "solid-state switch." To shut it off, power to the anode must be briefly interrupted or its voltage made negative. To convert a solid-state switch to a timing circuit it is necessary to add an RC timing network to the gate circuit (Figure 11*b*). When the capacitor is charged through resistor R_1 , the voltage builds up across C until it is sufficient to fire the SCR. Unfortunately, the voltage at which the SCR fires is dependent on ambient temperature; consequently, the usual practice is to modify the circuit to include a unijunction transistor that fires at a precise voltage independent of ambient temperature (Figure 11*c*). The voltage built up across a capacitor is a function of the magnitude of the applied voltage. If the voltage varies, the firing time of the circuit will change proportionally. Most good timers eliminate this problem by using a zener diode to provide voltage regulation. It will also suppress positive voltage transients that might damage components (Figure 11*d*). Very often, due to deficiencies in the

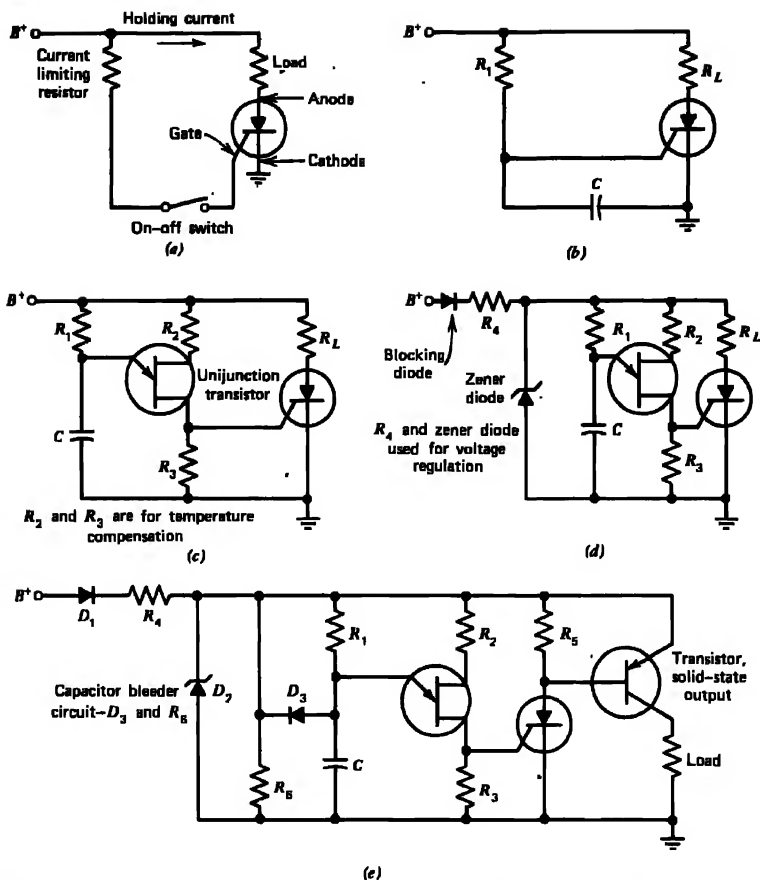


Figure 11. Development of an electronic timer. (a) Basic SCR circuit. (b) SCR circuit with RC timing network. (c) Unijunction transistor added. (d) Voltage regulator and transient suppressor added. (e) Complete timer, including solid-state output and bleeder. (Courtesy of McGraw-Hill Book Company. From Reference 3.)

power supply, negative voltage spikes are impressed on the circuit. These surges could turn off the SCR after it has become conductive. To eliminate the problem, a blocking diode is inserted in the circuit. Another valuable feature to look for is an additional transistor in series with the load to provide better impedance characteristics (Figure 11e). This is sometimes referred to as a solid-state output. Another feature worth looking for in a solid-state timer is a capacitor bleed circuit. A capacitor bleed circuit is a technique

for discharging the timing capacitor after the circuit has been deenergized. It consists of diode D_3 and resistor R_6 . The importance of this circuit is that it enables the timer to be recycled rapidly. If the timer is recycled faster than the capacitor can be discharged, the next timing period will be shorter than the specified time.

The basic problem associated with solid-state timers is the variation in resistance and capacitance as the temperature changes. The best capacitors to use for wide temperature range are teflon units. They hold a $\pm 1\%$ tolerance over a range of -60 to 80°C . Unfortunately, their cost is high and their size larger than competitive types. For a similar temperature range, polystyrene capacitors are a good second choice within a range of $\pm 2\%$. They are also expensive and relatively large. On the basis of cost and size, tantalum capacitors are used extensively by many timer manufacturers. Capacitance drops by about 12% at -55°C and also tends to vary with aging. The resistors used are generally wirewound with a temperature coefficient of $20 \text{ ppm}/^\circ\text{C}$.

So far we have discussed only timers that close a circuit at a certain time interval after the circuit has been energized. Some of the other modes are:

Delay on deenergization—applying power energizes the timer output. Removal of power initiates the time delay; the timer circuit output opens at the end of the period (Figure 12).

Delay on energizing—applying power energizes the timer output and initiates the time delay. At the end of the time delay, the timer circuit opens.

Pulse operating—same as the case above but designed to work with pulse circuitry.

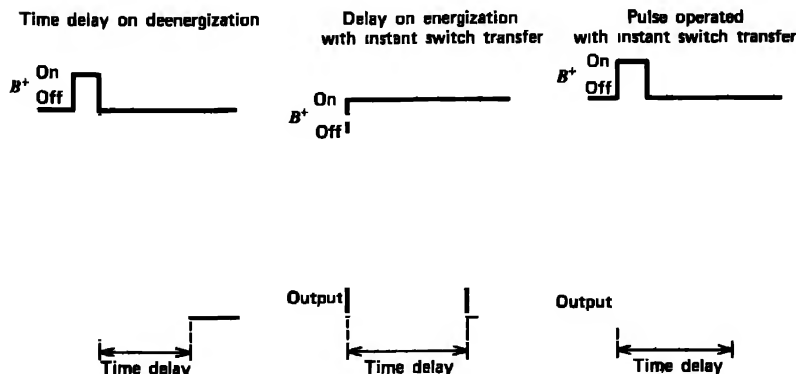


Figure 12. Basic timing arrangements. (Courtesy of McGraw-Hill Book Company. From Reference 3.)

The basic advantages of solid-state timers compared to other timers are as follows:

1. Extremely fast response time measured in microseconds, rather than in milliseconds as for electromechanical devices.
2. Reset time is also proportionally faster.
3. Elimination of contact bounce problems.
4. Excellent resistance to environmental effects.
5. Small size and low power consumption.

Disadvantages are:

1. Higher cost than for competitive units.
2. Cannot be adjusted as easily as electromechanical units.

4.7. INTEGRATORS

In counting digits, measuring spectrogram areas, simulating a process, or solving numerous other control problems integration is a basic operation. The integrating devices available are almost as varied as the applications. Integrators may be digital or analog, mechanical, electromechanical, electronic, or pneumatic.

Figure 13 outlines the ranges of accuracy, frequency response, and cost that can be expected for each generic type. These three characteristics are basic to integrator selection, but many others affect the choice. Some of the additional characteristics, as well as operating principles and limitations, are discussed below. Integrators that are no longer used in system work, such as the planimeter, have been omitted from the summary.

Counters are the most direct form of integrator, since they compile the total number of events that occur over a period of time. They may be actuated by mechanical, electromechanical, photoelectric, electronic, or pneumatic inputs. Because the counter is a digital form of integration, the error can be as low as 1 count in many thousands. Modern digital recording techniques coupled with fast-response counters can produce accuracies of 0.01% or better. Later examples will illustrate mechanical counters connected to other forms of integrators so that the readout is in digital form. Electronic counters are similarly connected to pulse-generating networks. In both cases the coupling device—gears, levers, potentiometers, transformers, or amplifiers—introduces a proportionality factor between total count and readout.

Electronic integrators are commonly used in the process industry as flow totalizers. The pulse generator in this case is a flowmeter. As the meter's blades rotate within a magnetic field at a speed proportional to fluid velocity, they generate a pulse train whose frequency is proportional to

Type	Accuracy (percent fullscale)	Frequency Response ^a (Hz)	Cost ^b
Mechanical counter	0.1 or better	0-100	Very low
Electronic counter	0.01 or better	0-10 ⁶	Moderate to Expensive
Ball-and-disk	0.1-0.5	0-20	Low
Dashpot	1.0-5.0	0-10	Very low
Gear train with escapement	0.1-1.0	0-10	Low
Electric motor	1.0-5.0	0-25	Low
Electric motor with tachometer	0.1-0.5	0-100	Moderate
Electric motor with mechanical filter	0.01-0.05	0-100	Moderate
Electric motor with drag cup	0.02-0.1	0-100	Moderate
Gyro	0.1 or better	0-5	Expensive
RC network	1.0-5.0	0-100,000	Very low
Operational amplifier	0.1 or better	0-10 ⁶	Moderate
IDVM	0.001 or better	0-10 ⁶	Expensive
Pneumatic	0.1-1.0	0-2	Moderate

^a Frequency response depends on such associated components as springs, amplifiers, and feedback loops. The figures shown are relative guides rather than absolute numbers.

^b Very low—under \$10; low—\$10 to \$100; moderate—\$101 to \$1000; expensive—above \$1000.

Figure 13. Summary of integrators. (Courtesy of Reuben H. Donnelly Corporation. From Reference 5.)

fluid flow rate. The pulses are simply counted to read out total flow, while a frequency detector produces an indication of instantaneous flow rate. This technique yields accuracies in the range of 0.1 to 1.0%.

Another electromechanical method of generating a frequency proportional to the process variable is a motor driving a multipole magnet adjacent to a magnetic reed switch. Figure 14 shows such a system applied to a belt weighing system, where the motor-switch-counter combination integrates the product of belt speed and loading to read the total material delivered.

4.8. MECHANICAL INTEGRATORS

The classic example of a mechanical integrator is the ball and disk mechanism. It consists of a pair of balls supported by a carriage and located

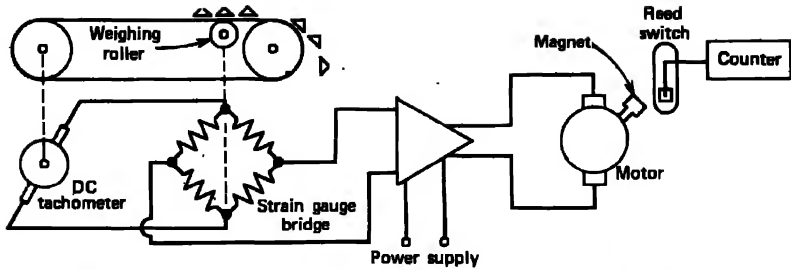


Figure 14. Pulse count integration. Acromag's motor-driven pulse generator drives a counter through a reed switch to integrate the amplified output of the strain gage bridge. The strain gage senses load and the bridge is excited by a belt speed signal. The counter therefore totals material flow. (Courtesy of Acromag, Inc.)

between a rotating disk and a roller or output cylinder (Figure 15). One input determines the rotational speed of the disk, and the other determines the position of the balls on the disk. Motion is transmitted from the disk through the balls to the output cylinder. The rotational velocity of the balls is directly proportional to their distance from the center of the disk. When they are positioned at the center of the disk, the velocity is theoretically zero; when they are at the periphery of the disk, the velocity is maximum. The output of the device is the angular position of the roller. The angular position, θ , is proportional to the integral of the product of the two inputs.

$$\theta = K \int \omega dR \quad (2)$$

where K = proportionality constant

ω = angular velocity of disk

R = linear position of the balls relative to the center of the disk

Both of the inputs may be variable or either one of them may be a constant. The most common input to the disk is a time function supplied by a constant-speed motor. The output in this case is the time integral of the input function R . If an acceleration transducer supplies the input to the ball carriage, the output is proportional to velocity. If a velocity transducer supplies the input, the output is proportional to distance.

The application shown in Figure 16 is an analog-to-digital integrator, where the ball and disk integrator sums the area under the analog curve on the chart recorder. Two forms of pulse generators—electromechanical and photoelectric—are shown on the output side of the integrator. Either type is available in a direction-sensitive design, so that the sense of rotation of the balls determines whether pulses are added or subtracted from the

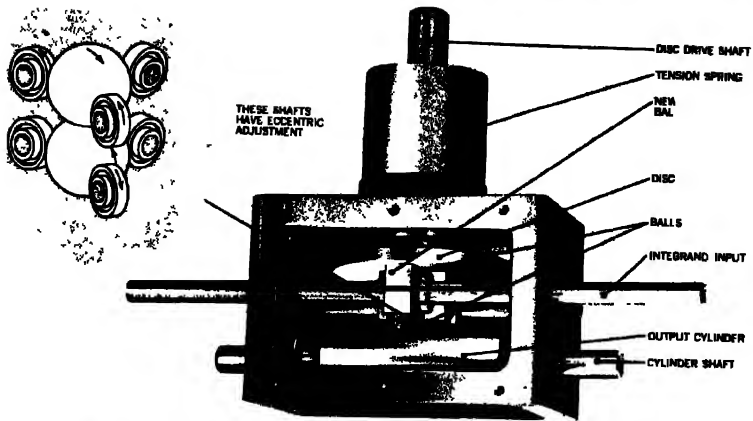


Figure 15. Ball and disk integrator. (Courtesy of Singer-General Precision, Inc. From Reference 6.)

count. Therefore, the counter reads out net area enclosed between the analog record and a baseline on the chart. The photoelectric pulser is capable of counts above 6000 per minute and is also available in a two-phase design suitable for actuating stepping motors and bidirectional counters.

In addition to integration functions, the ball and disk integrator may be used for many mathematical functions (Figure 17).

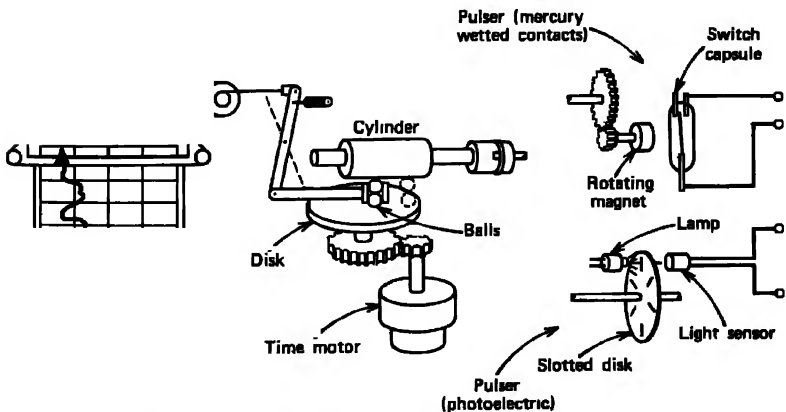
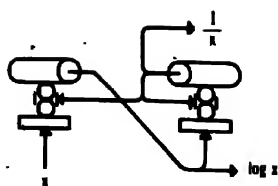
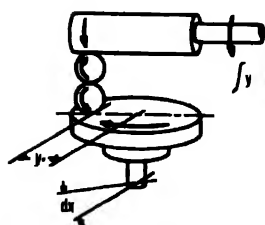


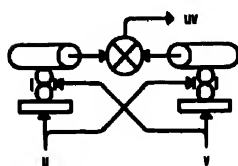
Figure 16. Area integrator. In Disc Instruments Inc.'s AD converter application shown here the balls are at the center of the integrator disk when the recorder pen position is on the baseline. By reversing its direction of rotation as the balls move across the center of the disk, the output roller responds to both positive and negative areas in the summation process. (Courtesy of Disc Instruments, Inc.)



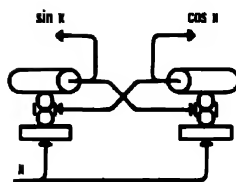
(a)



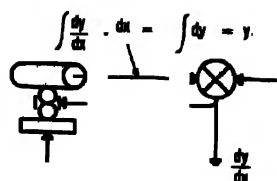
(b)



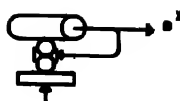
(c)



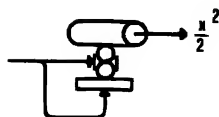
(d)



(e)



(f)



(g)

Figure 17. Computations performed by ball and disk integrators. (a) Computing the natural logarithm. The integral of dx/x equals the natural logarithm of x . The second integrator integrates the change of $\log x$, which is $1/x$. The quantity of $1/x$ is supplied as an input to the ball carriage of the first integrator, which computes a continuous value of the natural log of x . (b) Integration. In integrating $y dx$, the disk rotates proportionally to increments of dx , and the ball carriage is displaced from the center of the disk proportionally to y . The rotation of the output cylinder is proportional to the integral of y with respect to x . (c) Multiplication. Two integrators can be used to generate a product. The integral of $u dv$ plus $v du$ is equal to w . The two integrated quantities are aided by the gear differential. (d) Computing sine and cosine functions. The integral of $\sin x$ times dx is minus $\cos x$, and the integral of $\cos x$ times dx is $\sin x$. A simple feedback system employing two integrators functions as a resolver for obtaining the sin and cosin functions of x . (e) Computing rate of change of variables. To obtain dy/dx at the output of the differential, the difference between the inputs of the differential must be proportional to the rate of y input. If y changes at a given rate, dx/dy displaces the ball carriage of the integrator by a proportional amount. The displacement of the ball carriage occurs in such a way that y_1 , the roller output, has a rate of change equal to that of the y input to the differential but is displaced from y by an amount sufficient to maintain the ball carriage displacement. (f) Generating an exponential function. An exponential function can be generated by integrating e to x power times dx . Since this operation results in e to the x , a feedback circuit can be used to generate exponential functions. (g) Squaring a function. Since the integral of $x dx$ is $x^2/2$, an integrator can be used to square a quantity by supplying that quantity as an input, both to the input disk and to the ball carriage. (Courtesy of Singer-General Precision, Inc.)

Dashpots are generally used to provide a calibrated time delay, but in special situations they may be used as an integrator. The velocity switch, shown in Figure 18, is designed to close a circuit at a given velocity. It

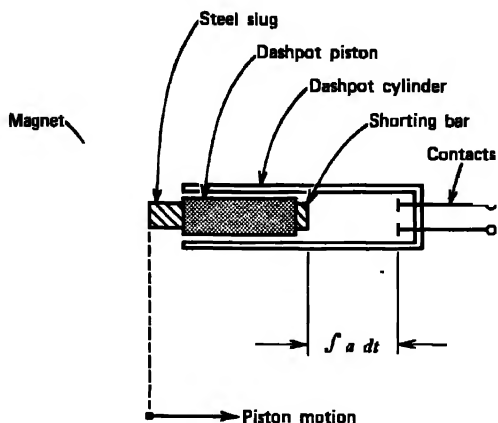


Figure 18. Velocity switch. The output motion of a velocity switch is the time integral of the applied acceleration. A more common use of this dashpot-based integrator is to provide a calibrated time delay. (Courtesy of Reuben H. Donnelley Corporation. From Reference 5.)

consists of a steel mass connected to a dashpot initially held in position by a magnet. When an applied acceleration produces a force on the mass that exceeds the holding force of the magnet, the dashpot piston starts to move down the cylinder. Its travel is restricted only by the retarding force of the air in the dashpot and the small effect of the magnetic field. Neglecting the magnetic field, the damping equation is as follows:

$$ma = D \frac{dx}{dt} \quad (3)$$

where F = force moving the piston

m = mass of the piston

a = acceleration of the piston

D = damping coefficient of the dashpot

dx/dt = piston velocity

Rearranging the equation and solving for x , the piston displacement becomes

$$= \frac{\pi}{D} \int_{t=0}^{t=t_1} a \, dt \quad (4)$$

This design has been used extensively in missile safety and arming mechanisms for actuating control circuits at a specified velocity.

One of the most inexpensive mechanical integrators is a gear train equipped with a runaway escapement. Before analyzing this device, it is necessary to review clocks, the parent device. Escapements used in timing devices are of two types: balance wheel escapements and runaway escapements. Balance wheel escapements are used in almost all clocks and watches and are accurate within $\pm 0.1\%$. Runaway escapements are not governed and produce timing accuracies within $\pm 7\%$.

Timing regulation of a clock is based on the mechanism shown in Figure 19. The balance wheel and hairspring form an oscillating system that rotates through a fixed angle in one direction and then rotates in the opposite direction through exactly the same arc. The period of oscillation is fixed and governs all subsequent events in the mechanism. The scape wheel has an intermittent rotation that counts the number of times the balance wheel oscillates and also supplies the balance wheel with the energy it loses in friction and windage. The power input to the scape wheel comes from the mainspring. Every complete oscillation of the balance wheel permits the scape wheel to advance one tooth. This motion is transmitted through gearing to the dial of the clock. Details of this governing action are shown in Figure 20.

A runaway escapement does not have a balance wheel. The energy stored in the mainspring dissipates itself as fast as the inertia of the mechanism permits; in short, the device attempts to run away when released. A typical design is shown in Figure 21. The first gear in the train is energized by a mainspring. Its motion is transmitted by the gear train to the scape wheel, which in turn allows the pallet to oscillate. The oscillating action of the pallet is the only restraint on the runaway action. In the example illustrated, the mechanism is used to integrate acceleration to obtain velocity.

The period of oscillation of the pallet may be approximated as follows:

$$t = \frac{1}{2} \sqrt{\frac{T_p}{2I_p \theta_1}} \quad (5)$$

where t = time for one oscillation (seconds)

T_p = torque transmitted by gearing to the pallet

I_p = inertia of the pallet

θ_1 = one-half the total angular rotation of the pallet

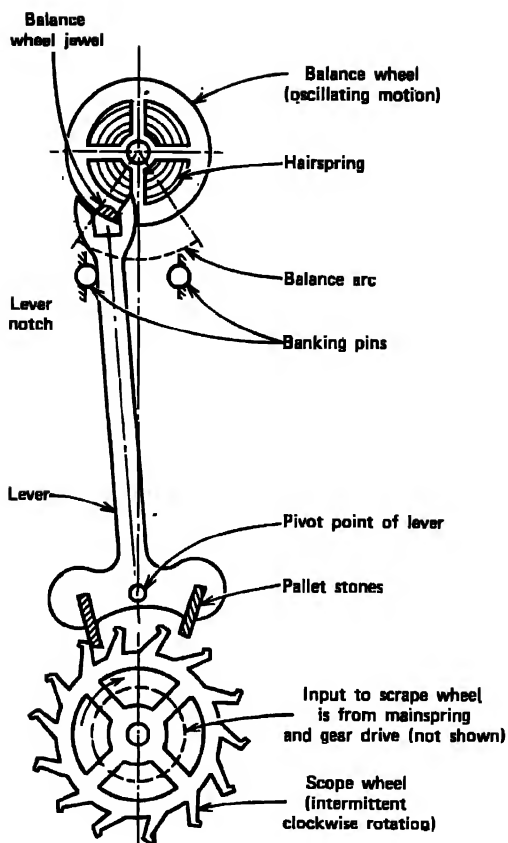


Figure 19. Principal elements of a balance-wheel escapement mechanism. Balance wheel and lever constitute a vibrating system with a fixed period, which governs the clockwise rotation of the scape wheel. (Courtesy of McGraw-Hill Book Company. From Reference 8.)

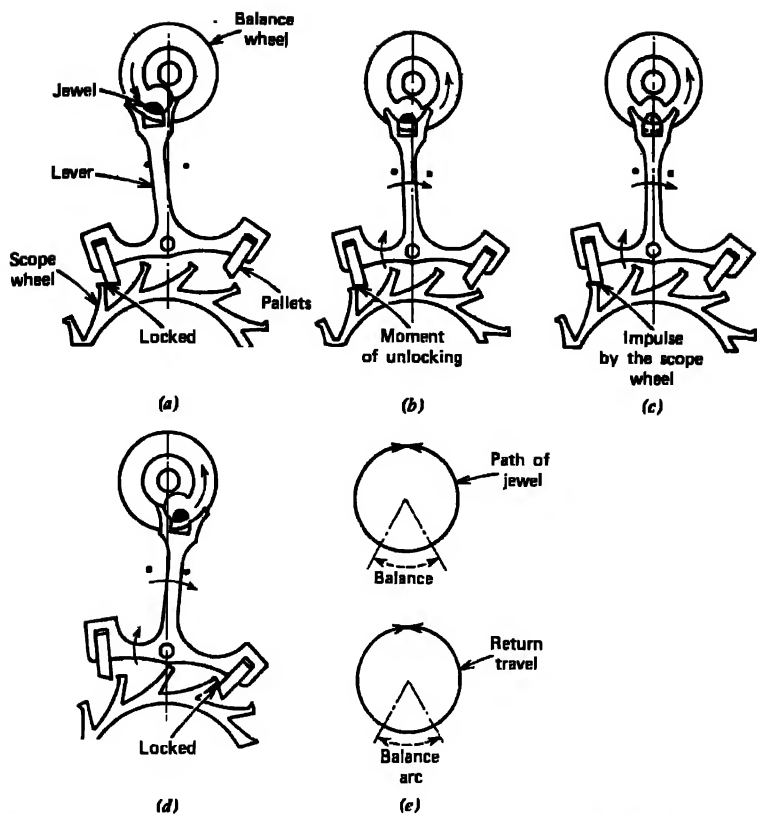


Figure 20. Sequence of events in a clock movement. (Courtesy of McGraw-Hill Book Company. From Reference 8.)

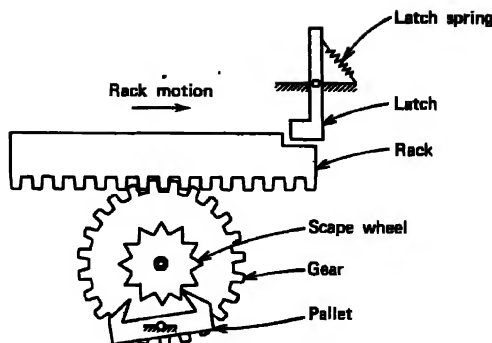


Figure 21. Runaway escapement. A runaway escapement provides a time base to integrate an acceleration applied to the rack in this gear train integrator. Position of the rack at any moment during its controlled travel is proportional to its velocity. (Courtesy of Reuben H. Donnelley Corporation. From Reference 5.)

This mechanism is used extensively in velocity and fusing devices. It is used for timing in the range of 0.05 to 300 sec and is remarkably resistant to the effects of shock, vibration, and acceleration. Most balance-wheel escapements cannot function in these types of environments. The chief virtues of runaway escapement are low cost and small package sizes.

4.9. ELECTROMECHANICAL INTEGRATORS

The heart of most electromechanical integrators is a motor whose output shaft position, θ , is the time integral of the power input,

$$\theta = K \int VI dt \quad (6)$$

where K = proportionality factor

VI = volts \times amperes or power input

t = time

The motor selected for use as an integrator must have a linear relationship between shaft position or speed and power input. The load applied to the motor is usually reasonably constant, and the problem is reduced to finding a motor with a linear speed-voltage characteristic. The simplest form of an electromechanical integrator consists of a motor coupled to a mechanical counter that counts shaft revolutions. Generally, a DC shunt motor or a two-phase AC motor is used. A DC or permanent-magnet shunt motor has a linear speed-voltage characteristic for constant torque. The two-phase motor characteristic is linear over a limited range; and integration will be accurate only within the specified range of speed and voltage. Another

necessary feature for integrator motors is some form of dynamic braking to prevent coasting when the motor is deenergized. If the motor rotor has substantial inertia, coasting can contribute large integration errors. A well-regulated power supply is also a necessity, since shaft speed is directly proportional to supply voltage.

Most processes are not "on-off" cycles but are, rather, proportional to demand. The technology used to integrate proportional processes is based on pulse width and pulse rate techniques. The easiest way to understand pulse-width integration is to consider the ball and disk integrator and its two inputs. Integration occurs only when both inputs are applied simultaneously. Pulse-width integration involves making one input proportional to the process being monitored while the second input is constant. The ratio of "on" time to "off" time is continuously proportional to the variable to be integrated. The first input is derived from a sensor that produces a mechanical displacement proportional to the process flow. It could be a flowmeter, pitot tube, thermostat, bourdon tube, or any other primary sensor. The output, if electrical, is converted to a mechanical displacement by an actuating mechanism and is then directed to the integrator. The second input to the integrator is from a synchronous motor. In this manner the input to the integrator is a series of pulses of variable amplitude and duration that are summed up as total flow or quantity. Variations on this theme are numerous; they are dictated by unique problems associated with the process, accuracy and price. One example is shown in Figure 22.

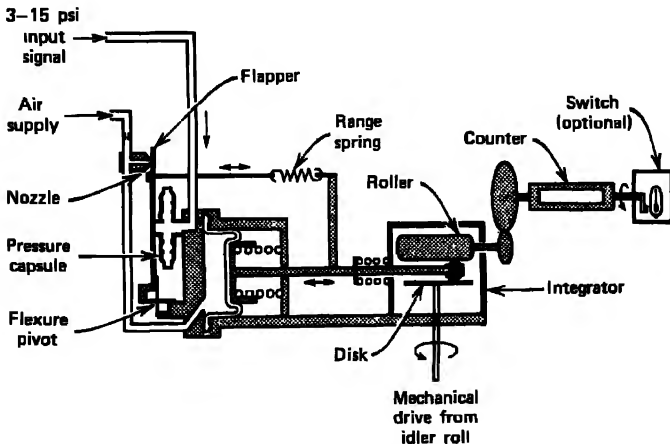


Figure 22. Pulse-width integrator. In Wallace & Tiernan's material weight totalizer, a pneumatic transmitter positions the carriage of a ball-and-disk integrator proportionally to the belt loading, and an idler roll on the conveyor belt drives the disk. (Courtesy of Wallace & Tiernan Corporation.)

Pulse-rate integration uses an electronic counter to add the total number of pulses generated by a sensor that converts process flow to a train of pulses. The sensor is typically an analog-to-digital device that produces a train of pulses whose frequency is proportional to the process being monitored (Reference 10). Typical integration accuracy for this device is 0.1 to 1.0%; pulse-width techniques are in the 1 to 2% range.

The most popular and successful method of improving the accuracy of motor-counter integrators is to provide some form of feedback in the system. The shaft velocity of the motor represents system velocity and the total number of revolutions recorded on the counter is the required integral. The tachometer, shown in Figure 23a, supplies a feedback voltage proportional to output shaft speed. The difference between the input voltage and the feedback voltage is amplified and drives the servomotor. The integration is accurate within $\pm 0.1\%$.

A simple motor-tachometer system has an inherent velocity error for ramp inputs. This error can be minimized by adding an integral control unit or mechanical filter to the system (Figure 23b). The mechanical filter is a very low-inertia motor with a torque output proportional to input voltage. The torque motor drives a low-noise induction potentiometer that energizes an amplifier, which in turn drives the servomotor. The voltage applied to the torquer is the difference between input voltage and tachometer feedback voltage. The mechanical filter is very heavily damped. The viscous damping is so much higher than friction and inertial effects that the system filters out most transients and is, in essence, a perfect integrator, capable of high gain and good frequency response. Accuracy is about $\pm 0.01\%$.

The drag cup integrator (Figure 23c) is a DC velocity servo. The torquer output, directly proportional to current input, drives a drag cup. The drag cup is an aluminum cup rotating in a magnetic field. As the cup cuts lines of flux set up by the magnetic field, eddy currents are induced in the cup that create a damping effect proportional to its speed. The summing junction shown in the servo diagram is actually the drag cup. The pickoff measures the rotation of the drag cup and energizes a motor that rotates the permanent magnet to produce a balancing force on the drag cup. At equilibrium, the counter torque generated by the magnet equals the torque exerted by the torquer, and the output shaft velocity is an accurate analog of the input current. Output shaft position is a measure of the integral of shaft velocity. Accuracies of $\pm 0.02\%$ have been achieved using this technique.

The integrating rate gyro is widely used for integration in navigational problems. This instrument senses a rate of turn about a given reference axis

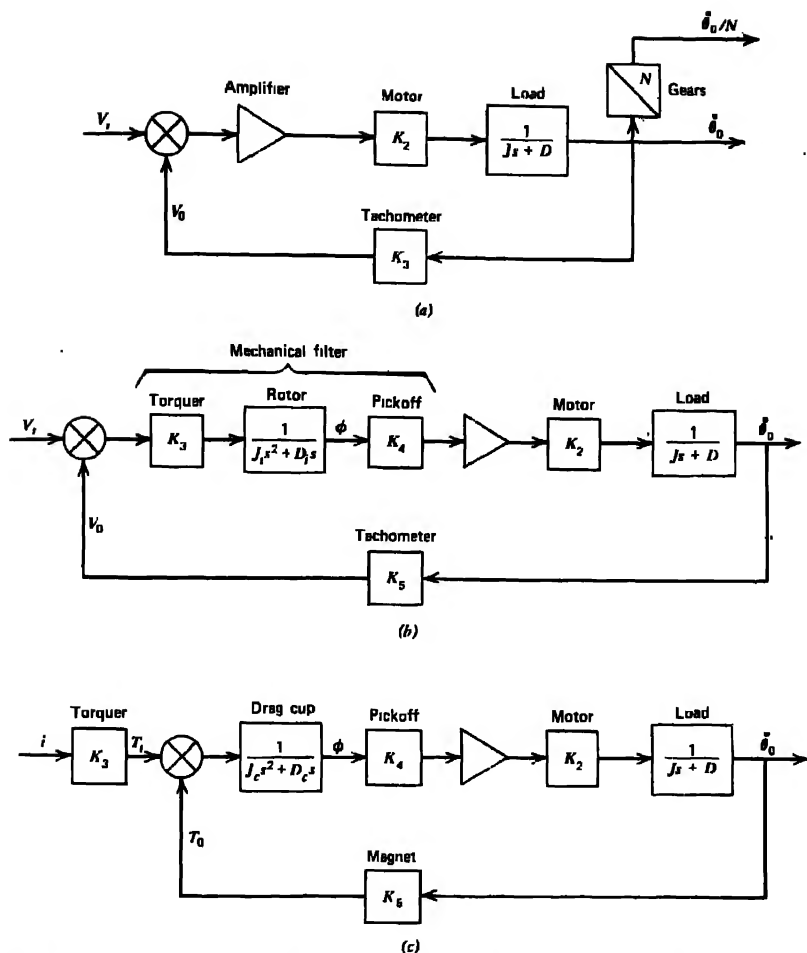


Figure 23. Servo integrators. Within certain limitations, the output shaft position of a motor is the time integral of its voltage input. However, accuracy and linearity are greatly improved by adding a feedback system to convert the motor to a closed-loop servo. Feedback may be through a simple tachometer (a), a tachometer with mechanical filter (b), or a magnetic field and drag cup (c). (Courtesy of John Wiley & Sons. From Reference 9.)

and produces an output voltage proportional to the integral of the rate. The essential parts of a gyro, (Figure 24) are as follows:

1. A floated cylinder containing a motor and flywheel to create angular momentum.
2. A supporting structure and case for the float.
3. A signal generator and torque motor to sense and correct for float rotation.

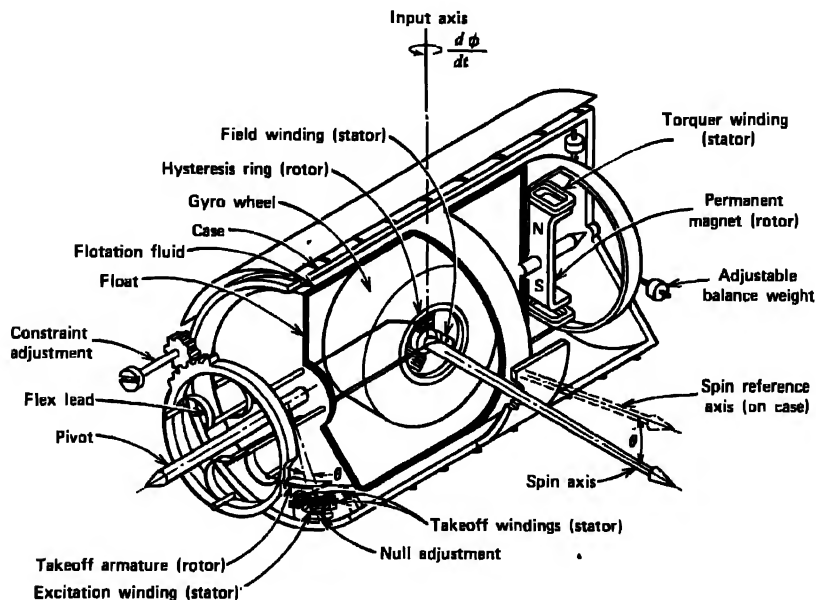


Figure 24. Gyro integrator. The integrating rate gyro is a null balance system. Input is an angular velocity about the precession axis, which causes an angular displacement about the output axis. The torquer current required to restore the initial position is proportional to the time integral of the input, or angular displacement. (Courtesy of John Wiley & Sons. From Reference 9.)

The gyro integrates angular rotation about its precession axis, which is the vertical axis in Figure 24. An input, $d\phi/dt$, about this vertical axis causes the float to rotate about the pivot or output axis by an amount θ . The signal generator measures the angular displacement, θ , and its amplified output drives the torquer to restore the float to its initial position relative to the case. This nulling technique prevents cross-axis sensitivity. The voltage drop produced by the torquer current flowing through a fixed resistor is

the gyro output and is the integral of the input rate. The explanation is as follows. A gyro is a second-order system, and the governing equation is

$$T = H \frac{d\phi}{dt} = J \frac{d^2\theta}{dt^2} + D \frac{d\theta}{dt} + K\theta \quad (7)$$

where T = the rotational torque

H = angular momentum of the gyro motor and flywheel

$d\phi/dt$ = the input rate

J = the moment of inertia of the float about its output axis

D = the viscous damping coefficient

K = a spring constant due to parasitic effects.

The gyro is designed so that K and J are extremely small, and the basic equation reduces to

$$\frac{d\phi}{dt} = \frac{D}{H} \frac{d\theta}{dt} \quad (8)$$

The displacement of the gyro float then is

$$\theta = \frac{H}{D} \int \frac{d\phi}{dt} dt \quad (9)$$

where H/D is defined as the gyro gain. The torquer current is proportional to θ and is therefore used as the integral of the input rate. The accuracy of gyros is measured in terms of drift rate. Good integrating gyros have drift rates of 0.001 to 0.01%/hour.

4.10. ELECTRONIC INTEGRATORS

The simplest electronic integrator is a resistor-capacitor network (Figure 25). When power is applied to the circuit, the output voltage across the capacitor, V_o , builds up as follows:

$$V_o = V(1 - e^{-t/RC}) \quad (10)$$

where V = applied voltage.

The circuit is a low-pass filter, and its effectiveness as an integrator depends on the relationship between its time constant, RC , and the time, t , in which the input signal undergoes an appreciable change. The charge and discharge curves for the capacitor in the circuit are exponential. If RC is much greater

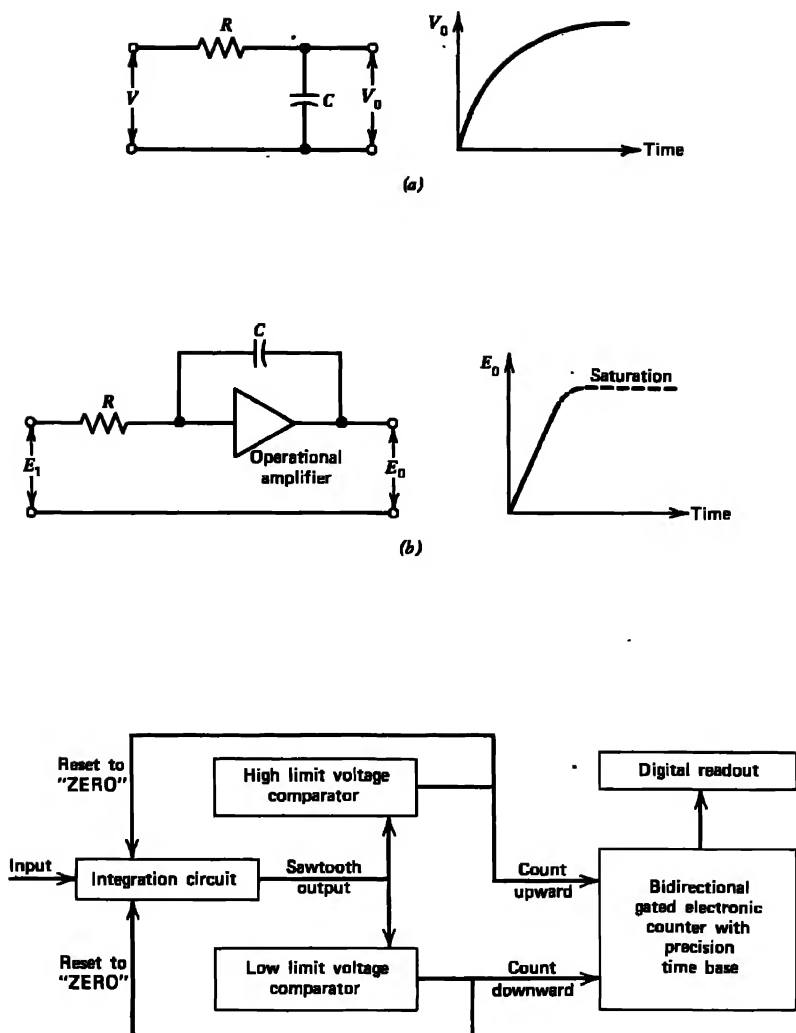


Figure 25. Electronic integrators. (a) Integrating accuracy of the simple RC circuit depends on its time constant and the input frequency. It provides short-term integration. (b) Operational amplifier with capacitor feedback integrates high-frequency signals most accurately. (c) Nonlinear System's integrating digital voltmeter responds to changes of polarity in the input signal, and has excellent noise-rejection capabilities. (Courtesy of Reuben H. Donnelley Corporation. From Reference 5.)

than t , the capacitor does not reach saturation, and the circuit behaves as an integrator. Under these conditions

$$= \frac{1}{RC} \int v \, dt \quad (11)$$

Although the RC network approximates an integration, it has several practical disadvantages. First, the charge leaks off the capacitor when held for any appreciable time. Second, the voltage build-up is limited to about 10% of the saturation value to prevent nonlinear effects. The most common application is as an input to a second circuit that functions when a prescribed voltage is attained. For example, it is used as the input to a relaxation oscillator in many timing circuits. The most commonly used electronic integrator is the operational amplifier circuit shown in Figure 25*b*. This is sometimes referred to as a current integrator. The operational amplifier is a high-gain, high-impedance circuit. It provides a charging current to the feedback capacitor equal to the current through the input resistor. In this way the input voltage, at the juncture of the capacitor and resistor, is held close to zero while the voltage across the capacitor is the time integral of the input voltage. The practical version of an electronic integrator is shown in Figure 25*c*. The integrating digital voltmeter (IDVM) combines an electronic integrating circuit with a counter. For a constant input, the integrator circuit generates a ramp function whose slope is proportional to input amplitude. When the output reaches a preset value, a comparator resets the integrator to zero and it starts to generate a new ramp. The IDVM generates a sawtooth pulse train whose repetition rate is proportional to input voltage. A bidirectional counter then counts the number of reset pulses that have occurred. The digital readout represents the integral of the input voltage with respect to time. A 100-mV input to the IDVM generates 100,000 pulses per second. The instrument can be equipped with built-in voltage attenuators and can cover the range of 1 μ V to 1200 V. The polarity of the input voltage determines the direction of the generated ramp function, so that two comparators and a bidirectional counter are necessary to provide true integration.

4.11. PNEUMATIC INTEGRATORS

Long before newer techniques were developed, pneumatic integrators were providing high standards of reliability and accuracy in the process industries. The totalizer shown in Figure 26 is typical of this instrument. An input signal in the 3 to 15 psi range, proportional to the square of the flow rate, drives the receiver bellows, which positions a flapper over a nozzle. As the pressure increases, the flapper approaches the nozzle, increasing the

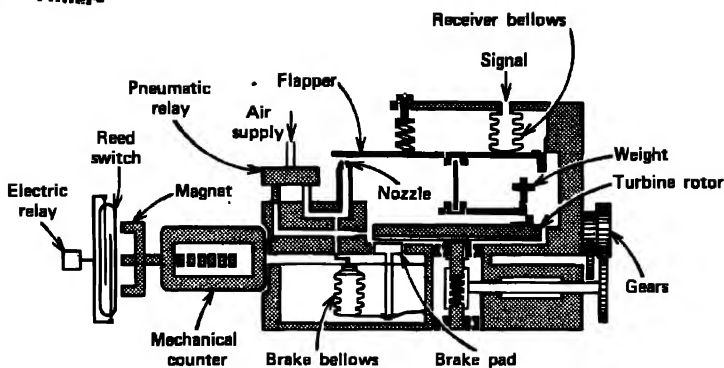


Figure 26. Pneumatic flow totalizer. In Foxboro's pneumatic flow totalizer, nozzle back pressure drives a rotor at a speed proportional to flow rate, and a mechanical counter accumulates the rotor revolutions. (Courtesy of Reuben H. Donnelley Corporation. From Reference 5.)

nozzle back pressure. The increased back pressure, acting through an air relay, regulates a jet that drives a turbine rotor. As the rotor spins, the centrifugal force on a weight attached to the rotor is transmitted through a mechanical linkage to the flapper to balance the force exerted by the input bellows. The force balance is maintained continuously. Since the centrifugal force is proportional to the square of rotor speed, and it balances the input signal which is proportional to the square of the flow rate, the speed of the rotor is directly proportional to flow rate. The rotor is connected through a gear train to a mechanical counter that totalizes the flow. In the instrument shown, an additional pulse-type output is provided. The output shaft is geared to one of the counterwheels and turns a four-pole permanent magnet. As the shaft rotates, the magnetic poles momentarily close the contacts on a reed switch. The reed switch operates a control relay to transmit a pulse to a remote pulse counter. The brake bellows shown at the bottom of the instrument drives the brake pad against the rotor when input pressure decreases suddenly. This damping action prevents spurious counts until the force balance is restored at the lower signal pressure. Pneumatic integrators frequently achieve accuracies better than 1% and in some cases approach 0.1%. Their most important requirement is a very clean supply of air.

4.12. SELECTION OF INTEGRATORS

The first step in selecting an integrator is to determine which generic type is suitable for the application. If no power is available, the choice is

limited to escapements, dashpots, and possibly ball-and-disk integrators. If high precision is required in a small package, the operational amplifier circuit and IDVM are indicated. When a reliable source of compressed air is available and frequency response requirements are modest, a pneumatic integrator may be the economical choice. Whatever the choice, there are two prime considerations for successful integration—drift and overshoot characteristics.

The generally accepted definition of drift is a change in output from the device without a change in input. In mechanical units this is caused by minute changes in dimensions due to stress relief and creep. Parts made of aluminum and magnesium are more likely to show this effect than units made of stable materials such as mild steel or brass. Preferred materials are beryllium copper and "NI-SPAN-C." Any system analysis should include a study of materials that may creep. All critical parts should be stress-relief-annealed to minimize this problem. Changes in dimension due to temperature variations and mismatches in expansion coefficients should also be included in system calculations. The only way of accurately determining drift effects is by actual testing in the system environment for extended periods of time. Drift is one of the most difficult mechanical problems to eliminate completely.

Electrical drift is also a headache. It is caused by heating and aging effects on component parameters. Most manufacturers of operational amplifiers will supply a carefully qualified figure on drift rate. For best results chopper stabilizers are preferred. Generalizations on operational amplifiers are difficult to make. Progress has been so rapid that today's problems will probably be tomorrow's triumph. An up-to-date set of manufacturers' catalogs is the only solution.

Drift is also affected by system noise. Units with inherently high damping, such as a dashpot, are less affected by noise than units with high frequency response.

The second major problem in integrators is overshooting. If the natural frequency of the integrator is about the same as the forcing function frequency, overshoots will occur. Braking systems, such as dynamic braking or pneumatically actuated pressure pads, are used to minimize this effect on electromechanical integrators. Electronic integrators are usually obtainable with high enough frequency response to eliminate the problem; however, they are much more susceptible to noise errors.

4.13 DIFFERENTIATORS

Certain applications require the use of differentiators (typically in the analog computer field, where a rate of change must be computed). Unlike

integrators, there are relatively few types of differentiators available. There are three principal methods:

1. Tachometric method.
2. The operational amplifier method.
3. The bellows method.

The tachometric method is illustrated in Figure 59 of Chapter 2. It involves the use of a complete servosystem and an operational amplifier. It is rather expensive and only moderately accurate.

The operational amplifier method shown in Figure 27 is one of the most common techniques. The RC network is selected to limit the effects of high-frequency noise which may be greater than the derivative output. Any high-gain operational amplifier is suitable for the circuit.

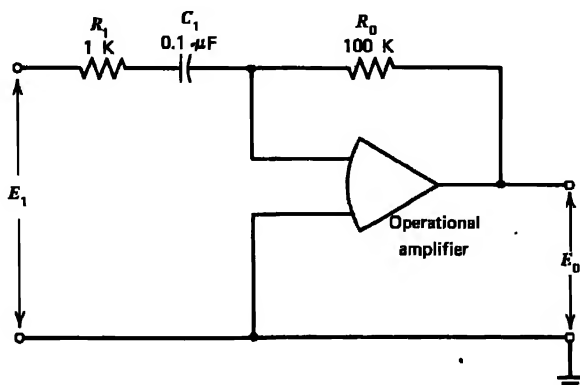


Figure 27. Electronic differentiator. $E_o = -R_o C_1 (dE_1/dt) = -1/100 (dE_1/dt)$. High-frequency cutoff $= 1/2\pi R_1 C_1 = 1.6$ kHz. Low-frequency cutoff $= 1/2\pi R_o C_1 = 16$ Hz. (Courtesy of Burr-Brown Research Corporation, Inc.)

The bellows method is illustrated by the aircraft rate-of-climb indicator in Figure 28. The cabin air pressure is supplied to a diaphragm within the case through a large tube and to the interior of the case through an orifice assembly and capillary tube. When a change of pressure occurs within the cabin, due to the aircraft climbing or descending, the change in air pressure appears within the diaphragm instantaneously. However, the pressure in the case does not change so quickly because of orifice action. A pressure differential exists between the diaphragm and the case and, consequently, the diaphragm expands or contracts.

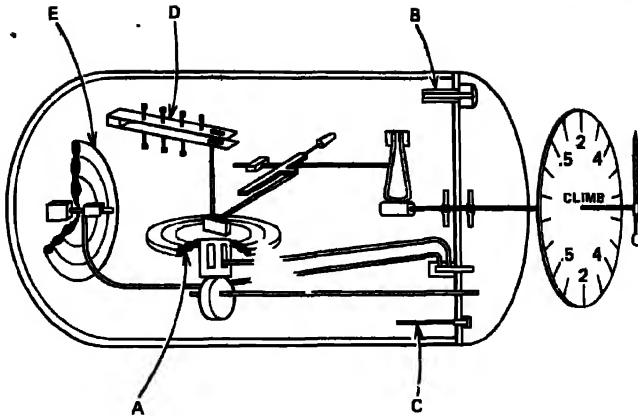


Figure 28. Rate of climb indicator. Changes in pressure due to change in altitude are transmitted quickly through a large connection tube to the inside of a diaphragm (A) and slowly through an orifice assembly and capillary (B and C) to the inside of the case. This creates a pressure differential, which causes the diaphragm to expand or contract proportionally to the rate of change of altitude. Adjustable restraining springs (D) are provided, which control the diaphragm deflection and permit accurate calibration. The action of the diaphragm is transmitted through a gear and leverage system to the pointer. As the plane assumes level flight, the pressure equalizes in the case and diaphragm and the pointer returns to zero. An overpressure diaphragm (E) prevents excessive speeds from damaging the mechanism. (Courtesy of Kollsman Instrument Company.)

The motion of the diaphragm is transmitted through a lever and gear system to position the pointer. This motion is proportional to the time rate of change of altitude. When the cabin altitude stabilizes, the pressure within the case becomes equal to that supplied to the diaphragm and the pointer returns to the zero position.

REFERENCES

1. How to Choose a Timer, *Product Engineering*, February 19, 1962, McGraw-Hill Book Company, New York.
2. J. Proctor, Timers, *Electromechanical Design*, May 1962, Benwill Publishing Company, Brookline, Mass.
3. E. Charkey, Solid State Timers, *Product Engineering*, August 30, 1965, McGraw-Hill Book Company, New York.
4. *Transistorized Industrial Timer Handbook*, Tempo Instrument Company, Plainview, N.Y., 1960.
5. E. Charkey, Signal Integrators, *Control Engineering*, April 1966, Reuben H. Donnelley Corporation, New York.

6. *Integrating Mechanisms, Section M-10*, Librascope Data Book, May 1961, Librascope Division, Singer-General Precision, Inc., Glendale, Calif.
7. G. W. Mechalec, Analog Computing Mechanisms, *Machine Design Magazine*, March 19, 1959, Penton Publishing Company, Cleveland, Ohio.
8. T. K. Steele, Clock Escapement Mechanisms, *Product Engineering*, January 1957, McGraw-Hill Book Company, New York.
9. I. Cochin, *Analysis and Design of the Gyroscope for Inertial Guidance*, John Wiley & Sons, Inc., New York, 1963.
10. D. E. Hoeschele, Jr., *Analog-to-Digital/Digital-to-Analog Conversion Techniques*, John Wiley & Sons, Inc., New York, 1968.
11. *Product Engineering*, April 1956, McGraw-Hill Book Company, New York.

Chapter V Relays, Stepper Switches, and Stepper Motors

Relays, stepper switches, and stepper motors are three stages in the development of modern control systems. Relay design is one of the oldest arts. It can be traced back to the classical experiments on electromagnetic coils, magnetism, telegraphy, and the telephone. The first modern relays were used for controlling motors. A relay is basically a switch that permits a small current to control a relatively high current circuit. It is the electro-mechanical "ancestor" of triodes and solid-state switches. Modern relay technology has been evolving for about 75 years, and every year has seen important changes and new classifications of products. The scope of relays available today is best understood by analyzing the categories of these devices.

5.1. CLASSIFICATIONS

Numerous attempts at classifying relays have been made by engineers, teachers, manufacturers, and associations. Some of the bases have been type of service, type of coils, number of contacts, type of latch, frequency response, physical size, type of energization, and number of steady-state positions. The definitive work so far has been published by the National Association of Relay Manufacturers (NARM). Relays are divided into 22 primary types and 35 special-purpose devices. The primary types are the following:

General-purpose	Power
Frequency-sensitive	Rotary
High-speed	Radio-frequency
High-voltage	Reed
Latching	Sensitive

214 Relays, Stepper Switches, and Stepper Motors

Mercury	Snap-action
Miniature	Subminiature
Motor starting	Stepping switches
Overload	Telephone
Plunger	Time-delay
Polarized	Vacuum

The special-purpose applications are the following

Antenna	Meter/indicating
Aerospace	Magnetic amplifier
Amplifier	Microwave
Adjustable set-point	Micro positioner
Audio-operated	Moving-coil
Coaxial	Microminiature
Clapper	Nanoampere
Chopper/modulator	Plug-in
Crossbar	Radiation-resistant
Differential	Solenoid-actuated
Date-programming	Surge-limiting
Electronic	Phase-sequence
Electromechanical	Solid-state
Explosionproof	Static
Fuse contactor	Thermal
Impulse-type	Voltage-comparing
Light-sensitive/photo	Voltage/current sensing
Low-level	

This is a formidable list to digest at one time, so let us start by discussing some of the basic parameters.

5.2. AC VERSUS DC RELAYS

DC relays are preferred to AC relays because they are smaller, more efficient, and less expensive for a given force requirement. AC units are power-limited because of the inductive nature of the circuit. The core of the coil must be fabricated from laminations to prevent eddy current losses, and shading coils must be provided to achieve a more uniform magnetic pull (Reference 4). Most designers presented with the problem of AC power source use diodes to rectify the current and then select conventional DC units. On certain home appliances where power consumption, size, and efficiency are not prime considerations, AC solenoids are still used. An

interesting form of DC relay is the polarized relay (Figure 1). It contains one permanent or electromagnetic magnet that provides an initial flux pattern; an auxiliary coil or coils provide a supplementary flux pattern that causes the armature to move in a prescribed direction. The supplementary flux can be arranged to add or subtract from the initial pattern according to the polarity, thus controlling the overall magnetic pattern and the position of the armature. The advantage of this technique is that a relatively small supplementary current is used to actuate the relay, resulting in faster response time. This technique is also used to actuate magnetic latching relays. The initial flux in the circuit may be insufficient to actuate the relay, but once the armature is moved, the initial flux is sufficient to hold it in the transferred position. This design is used extensively in reed relays.

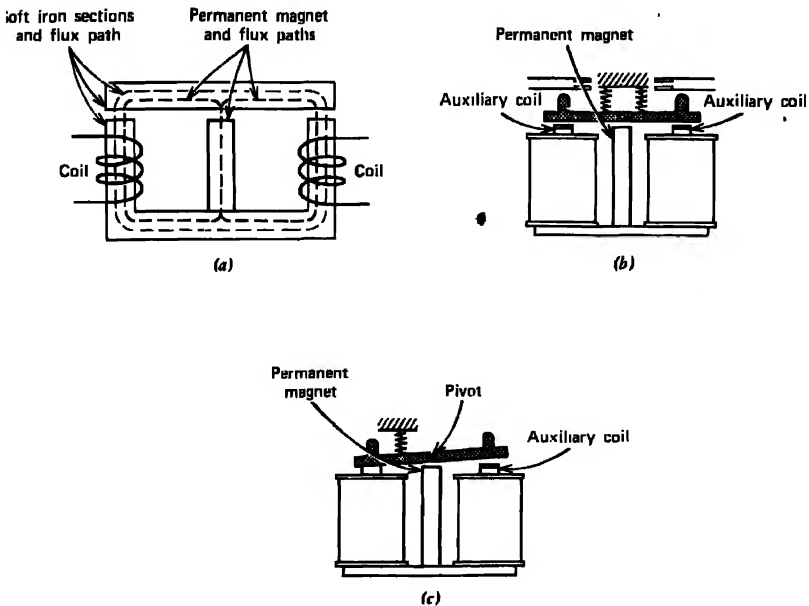


Figure 1. Polarized relay. (a) A basic magnetic circuit arrangement for a polarized actuator. (b) Polarized neutral center relay. (c) Polarized relay for pull-in and dropout operation. (Courtesy of Benwill Publishing Company. From Reference 18.)

The resonant reed relay is an example of an unusual AC relay. It contains an electromagnetic coil, designed so that its flux drives a vibrating reed. The reed is in essence an armature with a precise resonant or natural

frequency. When the frequency of the impressed voltage corresponds to the natural frequency of the reed, it will resonate. This causes an associated contact to close a circuit once during each cycle. The motion is intermittent but can be used for various control applications (Figure 2). One such

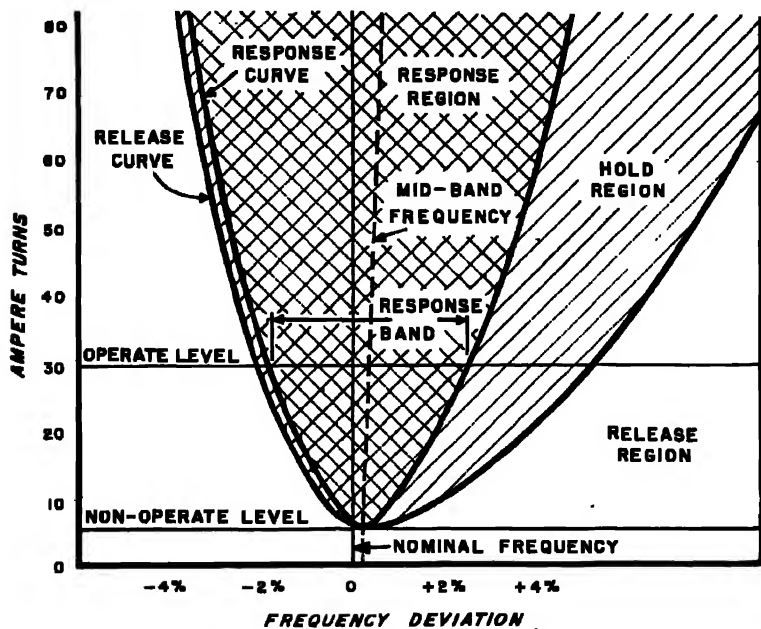


Figure 2. Resonant reed relay—frequency response characteristics. (Courtesy of James G. Biddle Company.)

system uses a series of reed relays whose natural frequencies range from 50 to 500 Hz, each connected to a specific control function. A low-frequency source is used to trigger each function by simply generating the proper frequency. In the range of 200 to 500 Hz it is possible to operate up to 16 channels with no interference between channels.

5.3. ELECTROMECHANICAL VERSUS SOLID-STATE RELAYS

Solid-state switches have replaced conventional relays in many switching applications. Although they excel where small size and response time are important, they are not a universal "cure-all." The simplest form of a

solid-state switch is a transistor or silicon-controlled rectifier; however, they are rarely used alone. Biasing resistors, shunting capacitors, and protective diodes are also required in varying degrees. In short, a solid-state switch is not a single component but a switching network. When discrete components are used, each has two or more solder joints that are potential failure points. Integrated circuits eliminate this restriction, but are limited to low power applications. On the basis of size alone, it is a toss-up between reed-type relays and solid-state switches, since reeds are a complete device requiring no auxiliaries.

When it comes to service life, semiconductors have no equal. They have no bearings to wear or contacts that pit and corrode. Response times are available down to nanoseconds compared to milliseconds for conventional relays. Resistance to shock and vibration is also superior. Electromechanical relays are preferable when electrical isolation between input and output circuits is important. Semiconductor leakage currents increase with time and cause further deterioration of circuit isolation. Conventional relays are also preferred when high transients or temperature variations are anticipated. Relays can be used for switching many circuits simultaneously; solid-state devices are inherently single-circuit elements. When high output voltages are required, relays are the proper choice. There is no important price difference between the two classes of devices (Figure 3).

	<i>Relay</i>	<i>Switch</i>
Response time	0.001–0.030 sec	$3 \times 10^{-9} - 1 \times 10^{-3}$ sec
Size	Small to large	Very small
Life	0.1 to 100×10^6 Hz (depending on load)	No limits
Circuit isolation	Very high	Low to medium
Power rating	Low to very high	Very low to medium
Resistance to transients	Good	Poor
Shock and vibration	25–50G	100G+
Temperature	Good at high temperatures; poor at low temperatures	Good at all temperatures, but upper limit restricted
Input signal	Operates with low to high inputs, AC or DC	Requires DC input and is amplitude sensitive

Figure 3. Electromechanical relay versus solid-state switch.

5.4. DEFINITIONS

To complete this discussion of some of the basic relay parameters, it is necessary to define the balance of the 22 primary types of relays listed by NARM.

GENERAL-PURPOSE RELAYS. This category is a "catch-all" for average relays that have no specialized function. Their chief application is to enable one circuit to actuate multiple circuits simultaneously. It may also be considered to be an electromechanical amplifier, since a relatively weak signal controls high-powered circuits.

FREQUENCY-SENSITIVE RELAYS. As mentioned in Section 5.2, they close a circuit intermittently when excited with a particular frequency. They are generally used in series with a second relay that latches and holds the circuit closed once the frequency-sensitive unit functions.

HIGH-SPEED RELAYS. They are units that operate in several milliseconds or less. Fast response time is achieved by keeping the mass and stroke of the armature to an absolute minimum. Polarized relays also are used for this purpose. Sometimes pulse techniques are employed to shorten response time. For example, assume that it is necessary to speed up the response of a general-purpose 28 V DC unit. The input switch is connected to an *RC* network that may be energized by as much as 300 to 400 V DC but discharges in about 5 msec. The power surge delivered to the relay coil is many times the rated value and causes the armature to move much faster than normally. However, since the power is available for only a brief period of time, the coil does not heat up excessively. This technique also is used extensively for speeding up solenoids in the computer industry. This method is safer than the utilization of a small armature travel, since this type of design makes it possible for the relay to function accidentally when subjected to shock or vibration.

HIGH-VOLTAGE RELAYS They are instruments that safely handle up to 10,000 V. Coil insulation is heavier than normal to accommodate the load safely. Extra precautions are taken in the design to provide an efficient thermal path between the coil and the heat sink member. These relays are used in high-potential cable testers and similar insulation testers, where they are clearly superior to semiconductors.

LATCHING RELAYS. Latching relays are designed so that the contacts stay in the transferred position after the unit is deenergized. This can be accomplished in a variety of ways.

1. By wiring the switch and coil in parallel with one set of contacts (Figure 4).

2. By providing a mechanical detent mechanism.

3. By using an auxiliary magnet to hold the armature in the transferred position. This is particularly useful in reed relays. Reset is accomplished by manual techniques or an auxiliary coil that works in reverse of the primary coil.

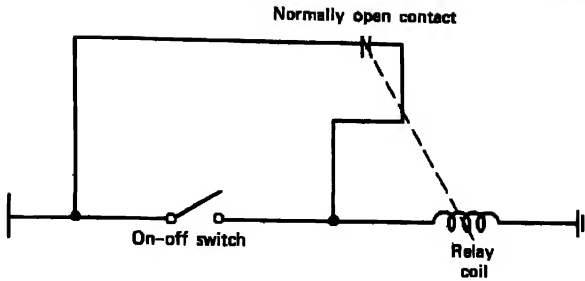


Figure 4. Latching relay circuit. Once the coil is energized, the contact closes and bypasses the switch.

MERCURY-WETTED CONTACT RELAYS. This component consists of a reed, in a close-fitting glass envelope, with its base immersed in a reservoir of mercury. The other end of the reed is positioned between two sets of fixed contacts. Capillary action causes the mercury to move up the narrow space between the reed and the inside of the glass envelope. This ensures that the extremity of the reed and the contacts are bathed in mercury. The advantages of this technique are elimination of contact bounce, long contact life, good current-carrying capacity, and freedom from contamination. Unfortunately, this device is position-sensitive; the flow of mercury is affected by the attitude of the capsule. There are variations of this component; one has a spring-biased armature and the other is biased by means of an external permanent magnet or coil. The spring-loaded design is capable of higher current ratings than the magnetic design. A great deal of effort is now being devoted to this design to make it smaller and less sensitive to position. It is the most promising competitor to solid-state relays in terms of life and size.

MINIATURE RELAYS. This is a qualitative designation used to distinguish the component from subminiature, medium, or large units. It should be considered to be a term that is more useful to buyers than to engineers.

MOTOR-STARTING/ARMATURE RELAY. This category is more a special-purpose application than a generic term. It is used to describe a dual-function relay. One section of the device is simply a power relay that connects the motor terminals to the line voltage. The other function of the relay is to open the motor-start winding circuit when the rotor approaches rated speed. This can be done by sensing either voltage or current. The voltage-sensing method consists of placing the second relay coil in parallel with the start winding; when the motor approaches rated speed, the voltage across the winding is sufficient to energize the relay coil, thereby opening the

contacts in series with the winding. When current sensing is preferred, the second relay coil is placed in series with the motor winding. The initial high current surge energizes the relay coil, closing a set of contacts in series with start winding. When the motor reaches its rated speed, the line current decreases, the relay coil is deenergized, and the contacts are in series with the starting winding open.

OVERLOAD RELAY. Overload devices are designed to monitor excessive current or voltage. When the danger criterion is reached, the relay actuates, latches, and energizes an alarm, siren, or warning indicator. Sometimes a time delay circuit is included to prevent false alarms, such as the "overloads" that normally occur when a motor is started.

PLUNGER RELAY. Solenoid-actuated relays are used when it is desirable to space the contacts far apart or when a large closing force is required. A special form of this device utilizes a solenoid-operated magnet to displace mercury so that it closes a series of contacts. The plunger, mercury, and contacts are hermetically sealed in an envelope. The actuating coil surrounds the envelope. The "Achilles' heel" is position sensitivity.

POLARIZED RELAYS. Previously discussed in Section 5.2, they are particularly useful for applications where the triggering signals are a series of high-speed pulses. They are also used when a bistable unit is required. They are available in most generic configurations.

POWER RELAYS. This term generally indicates capability to handle at least 15 A at 28 V DC or 115/230 V AC. Typical usage is in motor controls and power distribution panels. Many designs include circuits for protection of relay contacts during transients or "on-off" service.

ROTARY RELAYS. The chief use of rotary relays is where shock and vibration might cause conventional relays to actuate prematurely. The device has an armature that rotates to close a set of contacts. By balancing the armature, the instrument is virtually immune to normal forcing functions. Rotary relays are not so popular as conventional relays, which contain some form of spring to prevent accidental actuation. Unfortunately, the spring may be excited when its resonant frequency is reached; the rotary relay has no such problem.

RADIO-FREQUENCY RELAY. Radio-frequency relays are used to switch radio-frequency signals from one circuit to another. Design considerations include dielectric materials selected for low loss at radio frequencies, large contact gaps to minimize leakage, and high-voltage components.

DRY-REED RELAYS. Dry-reed relays (as opposed to mercury-wetted relays) consist of one or more pairs of flat reeds, separated by an air gap and

mounted in a hermetically sealed capsule. The capsule is surrounded by an electromagnetic coil that causes the reeds to deflect and touch each other when the coil is energized. Removal of the field allows the reed to separate under its own spring tension. Some designs utilize a mechanism that brings a permanent magnet into proximity with the capsule to actuate the reeds. Bistable or polarized reed relays are forms of this design. These are the smallest type of electromechanical relay available today. Ratings range from about 0.002 to 3.0 A. Response time is of the order of several milliseconds for the smallest units.

SENSITIVE RELAYS. This is a term reserved for relays that function with the smallest input signals. Typical ratings are 0.1 W or less. Most applications are in the precision instrument field.

SNAP-ACTION RELAYS. Snap-action devices are mechanisms that store energy without motion until a given increment of force is applied; the device then functions rapidly, utilizing all the previously stored energy. Numerous spring and magnetically controlled mechanisms perform this function. The electrical counterpart is a capacitive circuit. Snap-action relays are used for fast response and to minimize arcing during the contact transfer period.

SUBMINIATURE RELAY. This term is reserved for the smallest commercially available units, normally a unit less than 1 in. in length. The term micro-miniature has been used in the literature to indicate an even smaller unit. To date no formal agreement has been reached as to the range for each term.

STEPPING SWITCHES. A stepping switch is a coil-energized device that sequentially switches contacts in response to a series of pulses. Energizing the coil actuates a ratchet and pawl mechanism that moves a single pair of contacts across a series of stationary contacts. (See Section 5.9 for more detail.) At low voltage and power levels it has been replaced by solid-state switching networks. High voltage and power circuits still rely on stepper switches. They are also much more versatile than present solid-state equivalents.

TELEPHONE RELAYS. Telephone relays are one of the most popular and readily available multicontact units available today. They are characterized by a coil with a rather large length-to-diameter ratio. The contacts are also long and fastened to a bracket on the coil. Both AC and DC coils are available. An extremely large variety of contact configurations exist. Response time and power ratings are moderate. This relay was originally developed for the telephone industry.

TIME-DELAY RELAYS. The most common time-delay relay is one equipped with a series of shorting turns called a slug. This consists of several turns wound so that they produce a flux that opposes the flux generated by the main coil. This action delays the flux buildup initially and retards the collapse of the field when the coil is deenergized. The slug is physically a copper ring mounted on the core of the relay. The position on the core determines the time delay. Pull-in delays of 0.1 sec and dropout delays of 0.5 sec are practical. Telephone relays are most readily equipped with this type of device because of their long coils. This technique is effective only with DC relays. Other time-delay techniques are described in Chapter 4.

VACUUM RELAYS. This term applies to all relays with contacts in hermetically sealed envelopes. Colloquially this is also extended to envelopes filled with an inert gas. The big advantage is freedom from contact contamination and oxidation. Reed switches are the most common example of this component.

OPTOELECTRONIC RELAYS. This is a new category not previously listed in the NARM group. The relays consist of a light source* and photocell packaged in one subminiature enclosure. The relay functions when input signal energizes the light source, thereby illuminating one or more photoconductive cells, which act as a bistable switch. These devices are used when the ultimate in circuit isolation is required; values of $10^9 \Omega$ and 0.1 pF are achieved in practice. Size is comparable to the smallest relay of any other construction. Light-emitting diodes (LED) and mating silicon light sensors are also in this category (see Section 5.8).

5.5. RELAY CONSTRUCTION

The simplest form of an electromechanical relay consists of a coil, a frame to support the coil, an armature, a spring, and one or more pairs of contacts. When the coil is energized, a magnetic field is established between the armature and the coil. When the attractive force developed between the coil and the armature exceeds the holding force of the spring, the armature moves toward the coil and closes the contacts (Figure 5). Some of the terminology associated with the opening and closing of the contacts is shown in Figure 6. The bounce and chatter time of an electromechanical relay is one of its weak points; solid-state and mercury-wetted contacts do not have this problem. If the load is a high-inertia device such as a motor, the bounce will have little effect; if the load is a low-inertia device, it will follow each bounce.

* Present light sources include incandescent lamps, neon lamps, and light-emitting diodes.

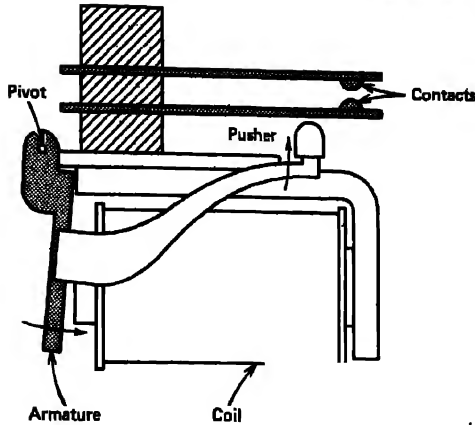
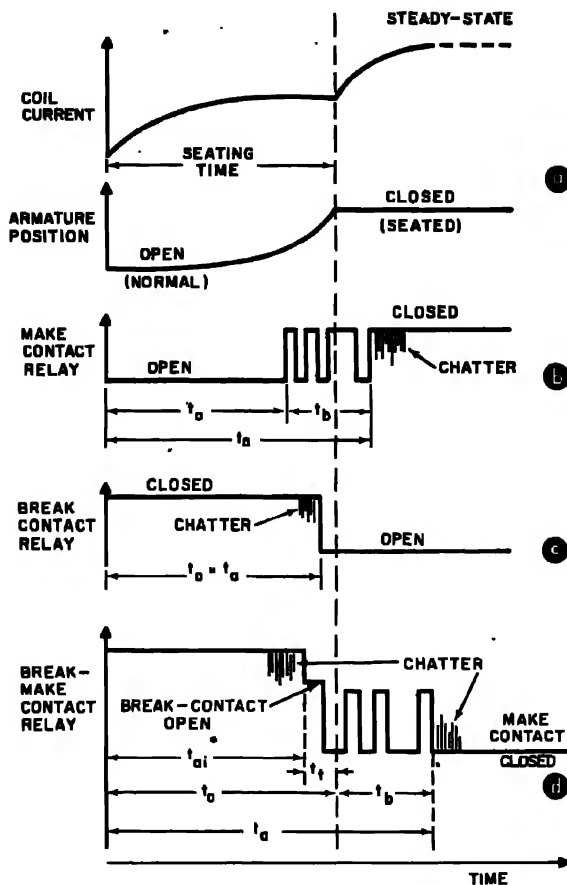


Figure 5. Relay construction. (Courtesy of Benwill Publishing Company. From Reference 18.)

The heart of a relay is the coil. As in all coils, the object is to pack as many ampere turns as possible into the available space. A detailed description of the coil design appears in *Electro-Magnetic Devices* by Roters (Reference 4). From a systems point of view the most important coil characteristics are its buildup and breakdown time constants and deviating characteristics. The ratio of coil inductance, L , to resistance, R , determines its time constant. The winding may be carefully "layer"-wound or "random"-wound with significant differences in inductances. Since the ratio L/R is a function of winding geometry and available power, it is best determined empirically for each unit. The heat transfer characteristics are more reasonably controlled by good design practices. The heat sink for relay coils is, primarily, the structure supporting the coil and, secondarily, the support for the relay structure. The most important consideration is getting the heat to the structure. Coil bobbins made of metals, such as brass, are much more efficient than plastic units. Metals conduct heat away faster and, consequently, require relatively little derating for sustained operation. Plastic bobbins are less expensive but retain the heat developed for longer periods of time. Another consideration is the type of wire used in the coil. Magnet wire is graded according to resistance to temperature; the common classifications are 105, 130, 155, 180, 200, and 220°C. The added protection afforded by 180°C wire is well worth looking for when selecting a relay. Cost differential is negligible. When the relay is not sealed in a hermetic envelope, the coil winding should be impregnated, or potted, with some



NOTE: 1. t_a = ACTUATION TIME
 2. t_b = BOUNCE TIME
 3. t_o = OPERATE TIME
 4. t_i = RELEASE TIME
 5. t_i = TRANSFER TIME
 6. t_{ai} = INITIAL ACTUATION TIME
 7. CHATTER ALSO KNOWN AS "GRASS" or "DYNAMIC RESISTANCE"

form of epoxy to prevent moisture from shorting adjacent turns. Most good relays are normally made in this way.

The comments about contacts in Chapter 4 are again applicable to relays. Most contacts are some form of silver alloy designed to function for millions of cycles. Further reliability can be achieved by an arc suppression circuit in parallel with the contact. One method is to provide an RC series circuit in parallel with the contact. The capacitor shunts the inrush current at the

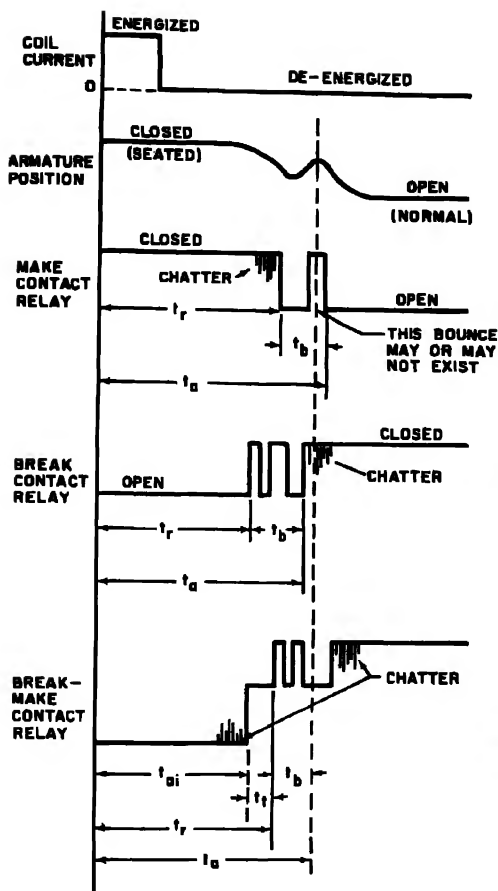


Figure 6. Relay terminology. (Left) Relay pickup function, with time as a base. Contact change is shown as a function of coil current and armature position (a). Note that armature position as well as coil impedance determines current. Transient phenomena for MAKE (b) and BREAK (c) contacts and combination MAKE and BREAK contacts in one relay (d) are depicted. (Right) Time traces of the relay dropout function. Coil current and armature position (a), MAKE contact (b), BREAK contact (c), and combined MAKE and BREAK contacts (d) are plotted as a function of elapsed time after the coil current has dropped to zero. Note that contact current flows as long as contacts are not open, independent of coil state. (Courtesy of Hayden Book Company. From Reference 5.)

moment of application, and the resistor limits the discharge current from the capacitor when the relay contact is closed (Figure 7a). When the contact is opened, a voltage is generated across the coil equal to $L(di/dt)$, where L = inductance of the coil and di/dt is the time rate of change of current. The magnitude of this voltage can be several times the supply voltage. The diode-resistor circuit shown in Figure 7b effectively limits it. The circuit

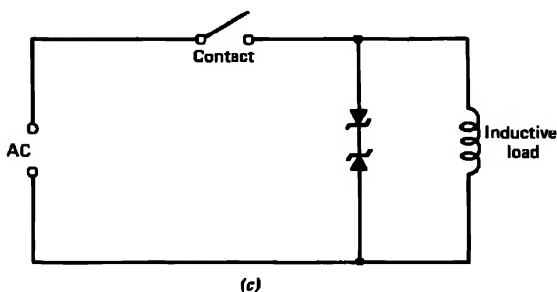
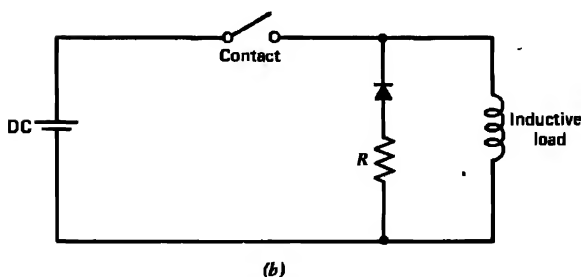
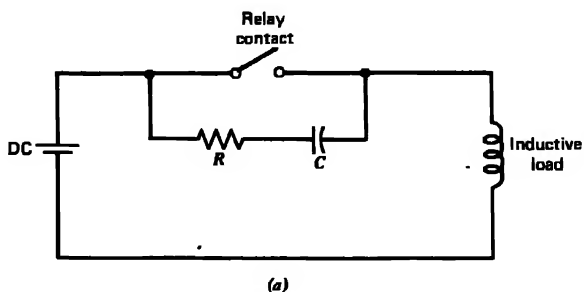


Figure 7. Contact protection.

shown in Figure 7c is for AC circuits. Numerous variations on these circuits are in use today.

Reed relays have a different variety of problem associated with their construction. Reed switches, or relays, consist of two flat, cantilevered reeds of ferromagnetic material surrounded by a dry, inert gas sealed in a glass envelope. A small gap separates the two overlapping ends of the reeds. When the coil surrounding the glass envelope is energized, it creates a magnetic field that causes the reeds to attract each other. The glass envelope is mounted in a resilient frame that absorbs much of the shock directed toward the switching element. A gasket supports the external reed leads so that they do not touch the frame of the assembly (Figure 8). The coil is

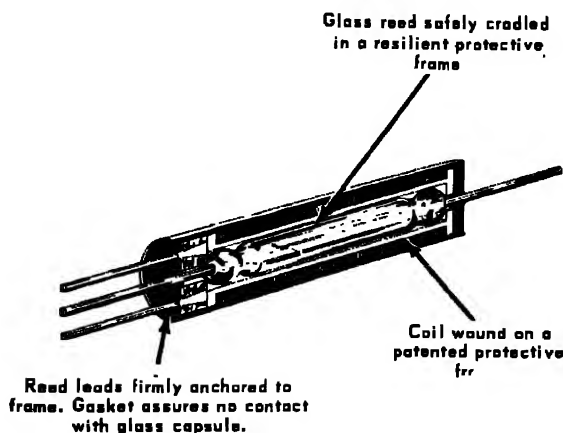


Figure 8. Reed relay. (Courtesy of Wheelock Signals Inc. From Reference 8.)

wound on the outside of the frame and utilizes it as a heat sink. The most important characteristic of reed relays is their fast response; a 1 to 20 msec response is characteristic of conventional hardware. These figures include contact bounce time. It is desirable to keep the natural frequency at 2000 Hz or higher, since the active element in this device is an undamped cantilever beam. When reed switches are used in military systems, they are normally expected to remain inoperative under vibration spectrums up to 2000 Hz; consequently, the natural frequency of the reeds is important. The contact resistance is normally held to 0.1 to 0.2 Ω at rated current. Mercury-wetted contacts provide even lower contact resistance and "bounceless" performance.

The third generic type of relay construction is the optoelectronic package. It is normally enclosed in hermetically sealed miniature cans, sometimes

called crystal cans, which are also used with conventional electromechanical relays. The lamp is normally a neon or subminiature incandescent bulb firmly mounted on one wall of the can (Figure 9). On either adjacent wall

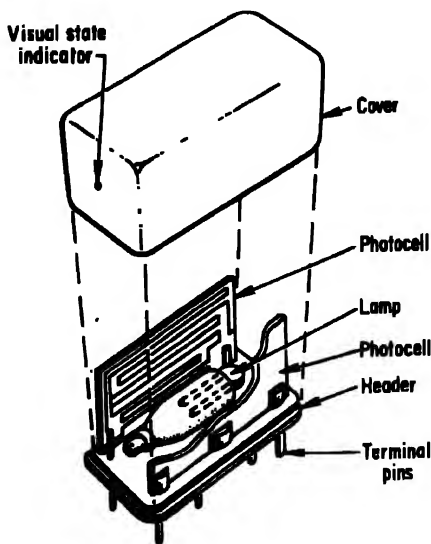


Figure 9. Construction of an optoelectronic relay. (Courtesy of Penton Publishing Company. From Reference 19.)

is a photocell. Each photocell corresponds to one independent pole of the output. As many as four poles are available in present designs. The basic limitation of the device is the relatively large impedance of the "output" circuit, which ranges from 500 to about 5000 Ω when turned on. This is undoubtedly a temporary condition in an expanding technology, since the photocell can be replaced with other types of solid-state photosensitive components such as phototransistors. The response time of neon bulbs is faster than that of incandescent bulbs, and they are used whenever relay response time is critical. Cushioning for bulbs is necessary to prevent damage during exposure to shock and vibration. Terminals are conventional hermetic-type headers. Solid-state versions, such as light-emitting diodes, have much smaller and more reliable constructions. They are structurally similar to an integrated circuit.

5.6. RELAY CIRCUITS

A vast amount of technology has been accumulated concerning relay circuits during the past 75 years. Each of the 22 primary relays has special qualities known best to the engineer who has actually built them. The 35 or more special-purpose relays also have individual traits. Rather than attempt the almost impossible task of covering all the combinations of the matrix, this discussion concentrates on only three generic types—the electromechanical, reed, and optoelectronic relays.

5.6.1. Electromechanical Circuits

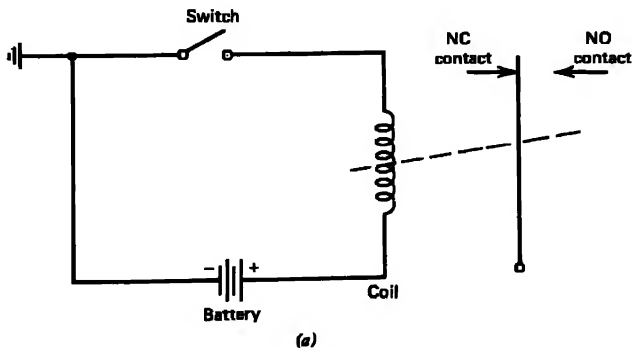
There are three principal reasons for employing a relay circuit:

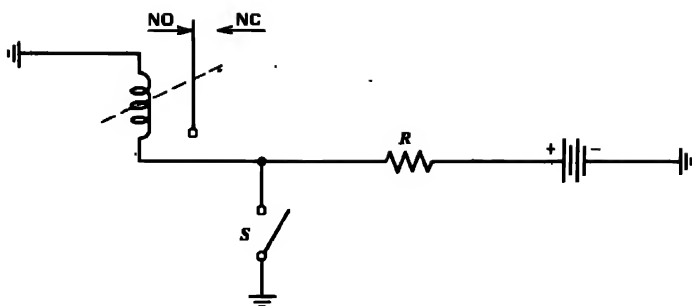
1. Power amplification—a low-powered circuit controlling the relay coil is used to actuate one or more sets of contacts that are in series with relatively high-powered circuits.
2. Numerical amplification—one circuit is used to control multipole configurations.
3. Special effects—frequency-sensitive circuits, time-delay circuits, latching circuits, and differential circuits are examples.

Some of the basic circuits used in achieving these results are discussed below (Figure 10).

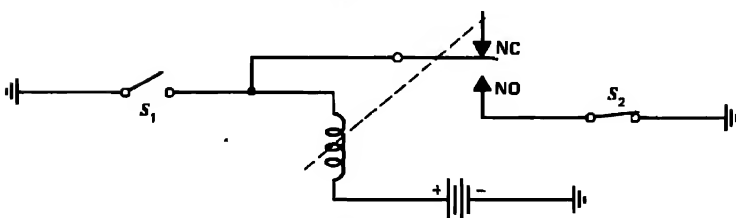
SERIES CIRCUIT. When the switch is closed, the coil is energized and the armature moves to touch the normally open contact (Figure 10a).

SHUNT CIRCUIT. When the switch, *S*, is open, the relay coil is energized and the normally open contact is closed. When the switch is closed, the

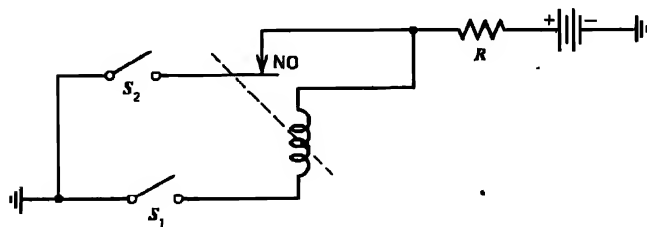




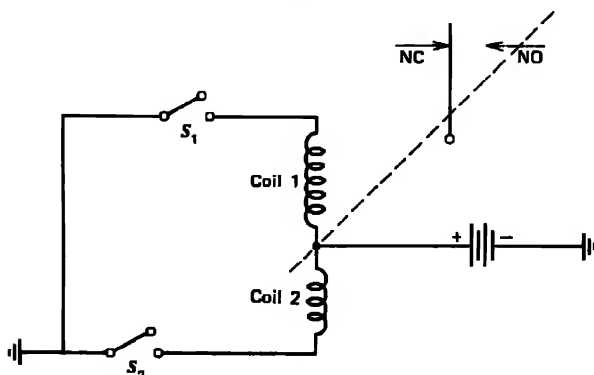
(b)



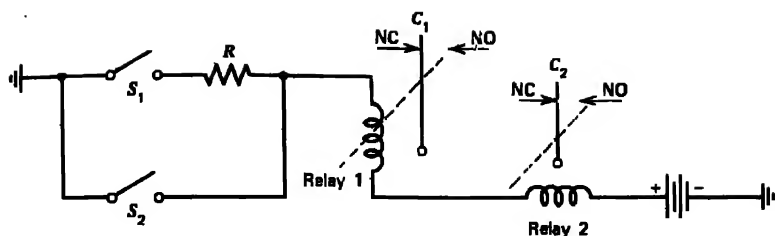
(c)



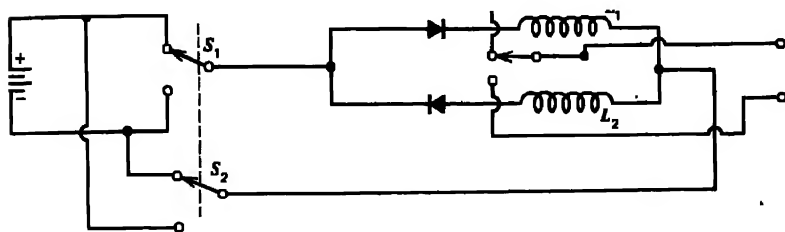
(d)



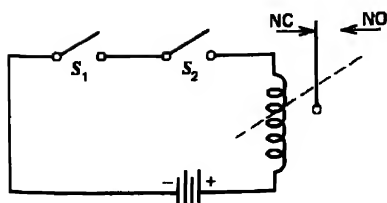
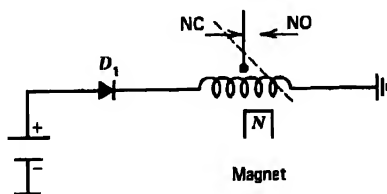
(e)



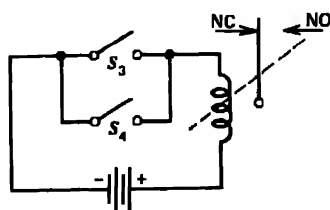
(f)



(g)



"And" circuit



"Or" circuit

Figure 10. Basic relay circuits. (a) Series circuit; (b) shunt circuit; (c) self-latching circuit; (d) lock-out circuit; (e) "AND" lock-out circuit; (f) double detector circuit; (g) polarized relay circuit; (h) magnetically biased polarized relay; (i) "AND" and "OR" circuits.

current is shunted away from the coil. This type of circuit always consumes power in resistor R and is more expensive than the series circuit (Figure 10b).

SELF-LATCHING CIRCUIT. This circuit is designed to remain in the actuated position once power is supplied momentarily. With S_2 closed, a momentary closure of S_1 energizes the relay coil closing the normally open contact. The coil remains energized so long as S_2 remains closed, regardless of the position of S_1 . This is the electromechanical counterpart of an SCR circuit where a momentary closure of the gate circuit causes the circuit to close and remain latched (Figure 10c).

LOCK-OUT CIRCUIT. This circuit prevents the solenoid from self-latching until an arming switch is actuated. While the arming switch, S_2 , is closed, it shorts the relay coil. When it is opened, and S_1 is closed, the relay is energized and its contact opens, decoupling S_2 from the circuit. The relay remains energized so long as S_1 remains closed. This circuit is commonly used in "interlock" systems (Figure 10d). The resistor, R , tends to slow relay response time.

THE "AND" LOCK-OUT CIRCUIT. The relay used is equipped with two coils wound so that their fluxes are in opposite directions. If both S_1 and S_2 are closed, the net flux generated is zero; if both S_1 and S_2 are open, the flux is zero. If either switch is closed, the relay will function. This circuit is used as a component in larger relay logic circuits (Figure 10e). Relays wound with two coils that produce opposing flux are called differential relays.

DOUBLE DETECTOR CIRCUIT. Sometimes it is necessary to detect two separate events that must occur in a prescribed sequence. The circuit for accomplishing this requires two relays. Assume that relay 1 requires a small current for operation and relay 2 requires a large current. When S_1 is closed, relay 1 is energized, but the current is insufficient to actuate relay 2. Closing S_2 shunts resistor R and a relatively large current energizes both relays. The contacts C_1 and C_2 monitor the events. Relay 1 must be specially designed to withstand the additional power (Figure 10f).

POLARIZED RELAY CIRCUITS. Defined broadly, a polarized relay circuit is one that indicates relative polarity of a signal. This is accomplished by using two coils that function only when a prescribed polarity is directed to each. The circuit of Figure 10g shows two ganged switches, S_1 and S_2 , that provide positive and negative input to the relay. Each coil differentiates the polarity of the signal with the aid of diodes D_1 and D_2 . The armature is pivoted with its fulcrum midway between the two coils and closes the corresponding contacts when each coil is energized. This is essentially a single-pole two-position device that indicates polarity by transferring a set

of contacts. Another commonly accepted definition of polarized relay is that of a device that is biased by an external magnet or coil so that only a small increment of flux is required to actuate it. This speeds up actuation time. The circuit is shown in Figure 10*h*. The chief disadvantage of this arrangement is susceptibility to accidental functioning under conditions of shock and vibration. Another variation of this design is to place the magnet so that it holds the contacts in the transferred position after the relay has been deenergized. The relay is returned to its initial position when a second coil is energized. It is also possible to obtain a unit that has two bistable positions corresponding to the polarity of each coil. This design employs two holding magnets.

THE "AND-OR" CIRCUITS. The ancestors of modern solid-state logic circuits are the simple circuits shown in Figure 10*i*. The "and" circuit will function when both S_1 and S_2 are closed. The "or" circuit will function when either S_3 or S_4 is closed.

5.6.2. Relay Race

One of the classic errors made in relay circuits is to assume that two relays with different response times will function in the anticipated time sequence. The usual result is that "Murphy's law" prevails and the wrong relay functions first. Energizing two relays simultaneously is known as relay racing. When using a relay circuit that involves a prescribed sequence of events, always wire the coil of the second relay in series with a contact controlled by the first relay (Figure 11).

5.6.3. Actuation and Release Times

The actuation time is the time interval from the application of power until the armature makes contact with the normally open contact. The contact bounce time is included in this figure. This is sometimes called

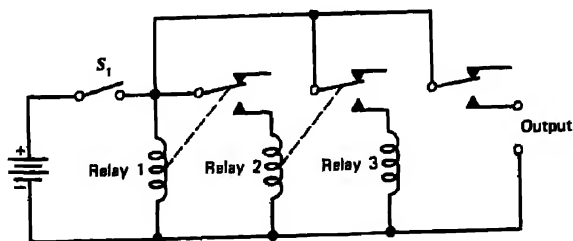


Figure 11. Three-stage relay sequencing circuit.

pull-in time. Release time is the interval between the time when power is removed and the time when the armature makes contact with the normally closed contact. Contact bounce is not included in this figure. Refer back to Figure 6. Temperature has a significant effect on actuation time; as the temperature increases, coil resistance increases, current decreases, force developed by the coil decreases, and acceleration of the armature decreases—the net result is slower response time. Release time is only a function of the decay rate of the magnet field and the spring constant of the armature; it is relatively unaffected by thermal effects. Contact bounce is the normal result of the impact of two spring mass systems moving at different velocities. It causes excessive contact wear due to arcing and can be completely eliminated only by using mercury-wetted contacts.

When it is necessary to speed up the response time, some method of temporarily applying greater than normal voltage to the relay coil must be used (Figure 12). The first method (Figure 12a) uses the normally closed

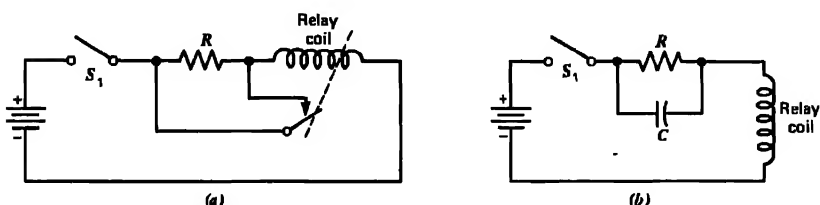


Figure 12. Fast-response relay circuits.

contact of the relay to shunt the resistor during the starting period, thereby allowing a higher than normal voltage across the coil. When the relay is actuated, the contact opens and normal voltage is applied to the relay. The second circuit (Figure 12b) uses a capacitor to bypass the resistor during the starting period. Fast release time is best achieved by selecting a relay with high spring constant on the armature. Delayed response is commonly achieved by winding a few turns of heavy copper wire in parallel with the coil so that their flux patterns are out of phase. This delays the buildup of the magnetic field. When it is desirable to delay the dropout of the relay, a capacitor is placed in parallel with the coil. When the switch is opened, the capacitor holds the coil in the actuated position until it is discharged (Figure 13). All relays are rated to function at nominal voltage, but the minimum actuation voltage should also be checked before completing circuit design. The holding voltage is another significant parameter.

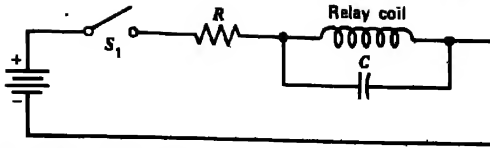


Figure 13. Slow dropout relay circuit.

5.6.4. Loading

A relay that is selected from a catalog with a given rating should be assumed to be for a resistive load. When the load is primarily inductive, such as a motor, the relay must be derated or specially designed for this type of service. Inductive loads normally require two to three times the normal resistive rated load and the contacts must be designed for this type of application. Selecting a heavy-duty relay for a small, inductive load is not a safe solution. The armature of a relay is essentially a cantilever beam with a concentrated load on the end; the concentrated load is the contact. Heavy-duty relays have relatively large contacts with the result that response time is slower than in small relays and contact bounce is worse. When inductive loads are required, some form of contact protection is essential. The circuits shown in Figure 7 are recommended strongly for these applications.

5.6.5. Radio Frequency Interference Suppression

One chronic problem in relays is RFI. It is caused by switching of energy across an air gap. The associated contact bounce produces an oscillating field that results in interference in the range of 0.1 to 30 MHz or more. The standard solution has been to minimize the arcing by conventional auxiliary circuits, such as those shown in Figure 7. Unfortunately, though helping the arcing problem, this does not always reduce RFI to acceptable levels. A practical solution developed for relays drawing less than 100 mA is shown in Figure 14a. One of the most effective techniques uses transistorized relay driver circuits. A relay driver circuit is employed when solid-state or integrated circuits serve as logic elements to determine when a relay is to be actuated. These elements are basically low-power units and must be coupled to a transistorized amplifier before they can be used with relays. Relay driver circuits are readily available for most relays (Figure 14b). The results obtained with these circuits offer the best solution to RFI problems thus far. Circuits for minimizing RFI are constantly being improved, and the solutions presented here are interim ones at best.

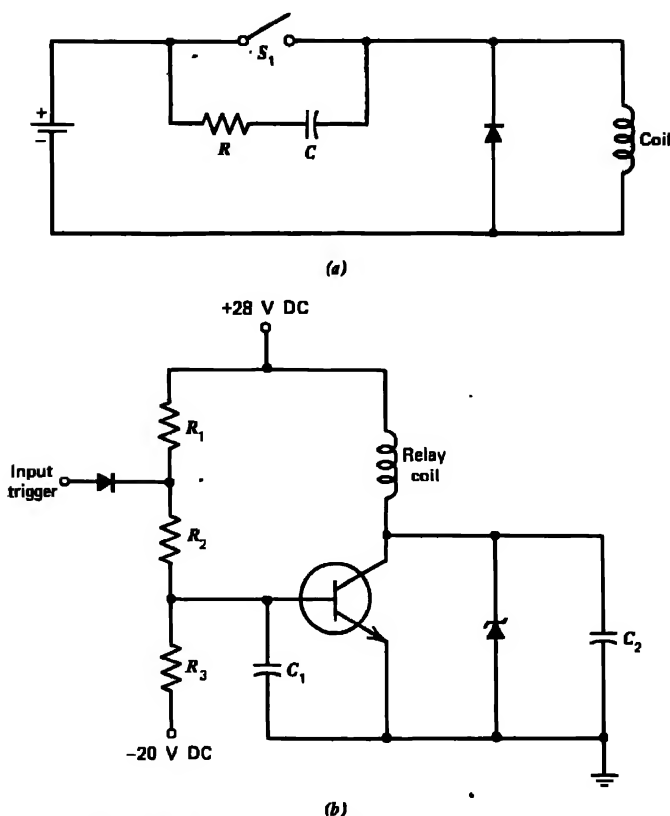


Figure 14. RFI suppression circuits. (a) Discrete component type; (b) solid-state type.

5.7. REED-RELAY CIRCUITS

When reed relays were first introduced, they were considered to be a possible replacement for electromechanical units when space and response time were the principal considerations. Through usage and evolution they have developed into more of a logic element, bridging the gap between relatively larger, slower electromechanical devices and smaller, more complex solid-state devices. The majority of new applications use reed relays as elements in a matrix, rather than one at a time. Reed relays are mechanically designed to fit into packages, containing six or twelve units each, that can be plugged into printed circuit boards or other receptacles

formerly reserved for solid-state or integrated circuits. Some of the advantages and limitations are as follows:

1. Insulation and circuit isolation is superior to those in most solid-state devices. Contact resistance is acceptably low.
2. Cost is low by modern (1971) standards—as little as \$1.50 per pole. This is competitive with solid-state units.
3. A reed, or electromechanical relay, is a working unit requiring no auxiliaries. Solid-state units require auxiliary resistors, diodes, and capacitors. Reliability is better than in complex circuits. They are therefore on a par with integrated circuits in this respect.
4. Life of a properly loaded reed switch runs into hundreds of millions of cycles and is therefore competitive with solid-state units.
5. Response time is clearly inferior to solid-state devices but is adequate for all but the fastest computers (Figure 15).
6. Contacts have resonant frequencies and contact bounce is also a problem.
7. Reeds have limited response to shock and vibration.

5.7.1. Types of Reed Relays

Relays are still in the evolutionary stage; some of the most popular relay types available today are listed below.

MOMENTARY CONTACT RELAY. When the coil is energized, the contacts close and remain closed so long as the coil is energized. When the coil is deenergized, the contacts open. This is the most common unit in use today.

PULSE REED RELAY. This type contains two independent coils; one actuates the switch and the other holds it in the closed position when energized. If the second coil is not energized, the relay performs as a normal momentary contact unit. The second coil does not have sufficient flux to actuate the relay. This design is useful for data storage. The information can be “erased” simply by deenergizing the second coil.

BISTABLE REED RELAY. This device contains two independent coils and a permanent magnet bias. A pulse to either coil causes the relay to function in the prescribed direction; the permanent magnetic bias then holds the reed in the transferred position until the opposing coil is pulsed. The switch is supplied normally as an SPDT unit. Like other electromechanical relays, reeds are available in the “break-before-make” and “make-before-break” configurations.

CROSS POINT OR MATRIX REED PACKAGES. This is a fascinating piece of hardware for mathematically oriented engineers. A group of relays are

	Crystal Can Relay	Miniature Reed Relay	Standard Reed Relay	Mercury Wetted Reed Relay	Switching Transistor Circuit
Operate time (including bounce)	5 msec	0.7 msec	2.0 msec	2.2 msec	< 10 μ sec
Operate bounce	50 μ sec	0.2 msec	0.2 msec	None	None
Release time	3 msec	0.06 msec	0.08 msec	0.08 msec	< 10 μ sec
Rated resistive load	2 A, 52 W	0.5 A, 10 W	1.0 A, 15 W	2 A, 50 W	< 0.5 W
Life @ rated load	< 1 $\times 10^6$ Hz	10 $\times 10^6$	2 $\times 10^6$	100 $\times 10^6$	Measured in hours of operation
Life @ dry circuit level	< 1 $\times 10^8$ Hz	1 $\times 10^8$ Hz	200 $\times 10^6$ Hz	> 1 $\times 10^9$	No
Suitable for inductive loads	Yes	Arc suppression required	Yes	No	No
Suitable for lamp loads	Yes	Light loads only	Yes	Yes	Yes
Suitable for motor loads	Yes	No	Yes	No	No
Resistance to voltage transients	High	Moderate	High	Low	Low
Contact isolation	High	High	High	High	Low
Closed circuit contact resistance (initial)	0.05 ohms	0.10 ohms	0.05 ohms	0.03 ohms	1.0 ohms
Open circuit resistance	1 $\times 10^8$ ohms	5 $\times 10^{10}$ ohms	1 $\times 10^{10}$ ohms	1 $\times 10^8$ ohms	1 $\times 10^6$ ohms
Coil power consumption	1 W	0.125 W	0.15 W	0.15 W	0.05 mW
Temperature stability	High	High	High	High	Low
Power supply regulation	$\pm 20\%$	$\pm 20\%$	$\pm 20\%$	$\pm 20\%$	$\pm 0.5\%$
Associated circuitry	Simple	Simple	Simple	Simple	Complex
Circuit reliability	Fair	Good	Good	Excellent	Excellent
Single pole CGT size and weight	Good	Excellent	Fair	Fair	Excellent
Resonant frequency	2 to 3 kHz	2 kHz	0.6 kHz	Low	No moving part
Vibration resistance (no contact openings)	20 to 30 g	30 g	15 g	Low	No moving part
Shock resistance	50 to 100 g	50 g +	35 g +	Low	No moving part
Trouble shooting	Easy	Easy	Easy	Easy	Complex
Circuit cost (approx.)	\$8	\$3	\$3	\$5	\$10

Figure 15. Relative merits of reed relays. (Courtesy of Wheelock Signals, Inc. From Reference 8.)

packaged together to form a matrix. Each relay is equipped with two coils; one a holding coil. When the two coils are energized, sufficient flux is generated to actuate a particular reed in the matrix. The holding coil then keeps the element closed. The unit is opened when one of the coils is reverse-pulsed. This unit is somewhat of a specialty, but is available from several manufacturers.

LOGIC RELAYS. Several manufacturers have packaged "and" and "or" configurations of relays in one package. The "and" configuration requires the application of two inputs simultaneously; the "or" package functions when either of two inputs is present. These relays are direct competitors to the solid-state packages currently available.

Some of the packages feature relays with magnetic shields around them to prevent magnetic interaction with adjacent units.

5.7.2. Dynamic Characteristics

The time between the first application of power and the initial movement of the reed is defined as the coil buildup time (Figure 16). It is a function

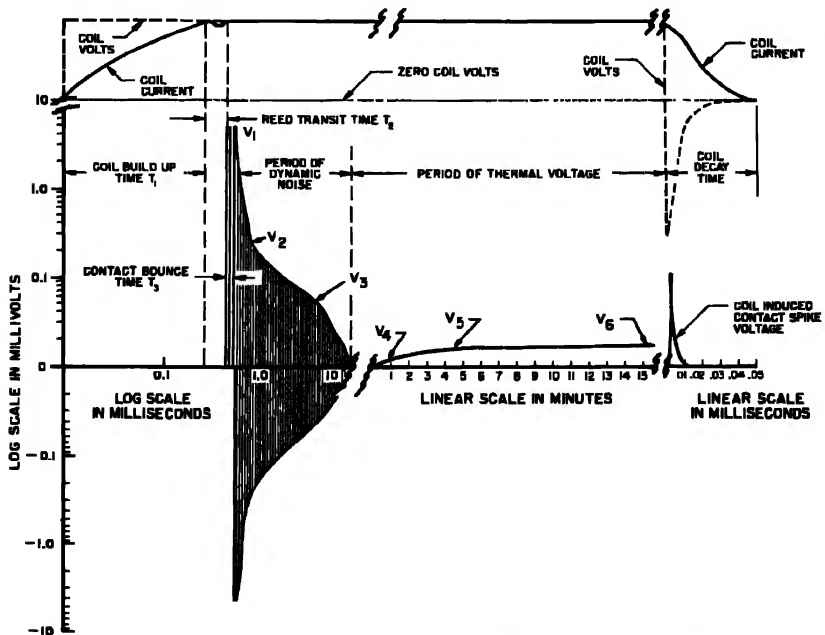


Figure 16. Reed relay dynamic characteristics. (Courtesy of Wheelock Signals, Inc. From Reference 8.)

of the L/R ratio for the coil and the power applied. The next interval, called the reed transit time, is the time required for the reed to swing through the air gap. The reeds are driven together with such force that they normally strike one another and bounce several times; this is called the period of contact bounce. It can be reduced by lowering the coil power or using mercury-wetted contacts. Dynamic noise is caused largely by a magnetostrictive voltage generated by the stressed reed in the field of the coil. It is an AC signal in the high audio range that decays as a damped sine wave. The last dynamic stage is a relatively long period during which a DC voltage is generated by the reeds as they are heated by the coil. The magnitude of this voltage is small and fairly constant. The instant the circuit to the relay coil is opened, a coil spike, or counter EMF, is generated that is a function of the L/R ratio of the coil. It can be minimized by conventional arc suppression circuits. The coil decay time and release time are a function of the coil design, magnetic circuit, and spring constant of the reed. Figure 17 shows representative values of these parameters.

5.7.3. Mercury-Wetted Contacts

Reed relays, like their electromechanical counterparts, are available with mercury-wetted contacts. The chief advantage is the elimination of contact bounce. The units are designed so that the reeds are covered with a thin layer of mercury due to the capillary action between the reed and the surrounding cylinder. The limitation of most of these relays is that they are position-sensitive; they are designed to be used in the horizontal plane. During the past five years, several companies have developed units not sensitive to gravity (Figure 18). The switching element is a permanent-magnet cylinder rather than a reed. It is enclosed in a hermetically sealed glass tube with a contact at each end and two coils. The clearance between the cylinder and the glass wall is held to a very small value and, as a result, the capillary forces are very high. Both contacts are covered with mercury. When the coil is energized, the switching element moves down the tube and magnetically latches to the contact until the second coil is energized. Actuation time is slower than for comparable dry reeds, 2.5 msec, but the unit can be mounted in any position and behaves well under shock and vibration environments. Typical units are 0.040 in. dia. \times 0.25 in. long.

5.7.4. Loading

The ideal load for most high-speed reed relays is a low-level, constant-voltage one. The application of transient voltages sufficient to cause arcing and welding are the chief reason for short contact life. Incandescent lamp

	T ₁	T ₂	T ₃	V ₁	V ₂	V ₃	V ₄	V ₅	V ₆
				Peak-to- Peak Dynamic	Peak-to- Peak Dynamic	Peak-to- Peak Dynamic	Thermal EMF @ 25°C	Thermal EMF @ 25°C	Thermal EMF @ 25°C
	Coil Buildup Time (msec)	Reed Transit Time (msec)	Contact Bounce Time (msec)	Noise Initial (mV)	Noise 1 msec (mV)	Noise after 5 msec (mV)	after 1 min (μV)	5 min (μV)	after 15 min (μV)
Series 372	0.2	0.1	0.1	3.0	0.6	—	8	18	12
Series 330	0.4	0.2	0.2	2.0	0.3	—	1	5	6
Series 3421	0.4	0.2	0.3	2.0	0.3	—	0.5	1.3	1.6
Series 262-1A	0.6	0.3	0.3	2.0	0.3	—	1	5	6
Series 262-4A	0.6	0.3	0.3	2.0	0.3	—	4	7	6
Series 3002-1A	2.0	0.8	0.3	3.0	0.6	—	1	4	4
Series 3002-1A (mercury)	2.0	0.8	0	2.0	0.5	—	—	—	—
Series 3002-1C	2.0	0.8	1.5	3.0	0.6	—	2	7	3
Series 3002-4A	2.0	0.8	0.3	3.0	0.6	—	4	8	10

Figure 17. Reed relay design constants. (Courtesy of Wheelock Signals, Inc., From Reference 8.)

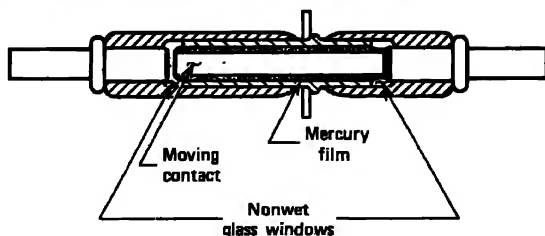


Figure 18. Mercury film switch. (Courtesy of Fifth Dimension, Inc.)

loads are most likely to produce this effect, since the initial cold resistance may be one tenth the normal operating resistance. Large reed relays or mercury-wetted relays are recommended for this type of application. Large capacitive loads requiring a large initial current surge will produce a similar result. Inductive loads, normally requiring two to five times rated current, may be employed with reed ratings by using arc suppression circuits. Motor applications should definitely be equipped with oversized, high-current-rated reeds. These are suitable for only very small motor applications.

5.8. OPTOELECTRONIC RELAYS

These are a new generic form of relay that is still in the early stages of evolution. Nevertheless, they are now on the market as shelf items and are in limited use. In contrast to conventional electromechanical or reed relays, they do not have an output that is essentially a switch closure. Photoconductive cells have an output characteristic that is more like a variable resistor than a switch. When no light reaches it, typical resistance values are in the range of 1 to 10 M Ω ; as the light increases, this value declines to the range of 100 to 1000 Ω . It is therefore not a replacement for a relay. The circuitry it controls must be specially designed to function with this device. The dynamic resistance range, 10^7 to 10^3 Ω , must be used to control some secondary circuit. This design approach is similar to that used in other photosensitive devices such as the photochopper; the principal difference is that optoelectronic relays have the light source enclosed in the same package as the detector, where it can be precisely controlled. Most of the conventional single-pole circuits synthesized with conventional relays can be designed to use these devices.

Another form of this device is called an optically coupled isolator. The light source is a solid-state device—a gallium arsenide diode. The detector

is a phototransistor. The advantage of this arrangement is virtually unlimited life, very high resistance to shock and vibration, and small envelope. The output is characteristically like any other transistor; it is proportional to the voltage applied to the input diode within the specified range of the device (Figure 19). The logical extension of the optically coupled isolator is

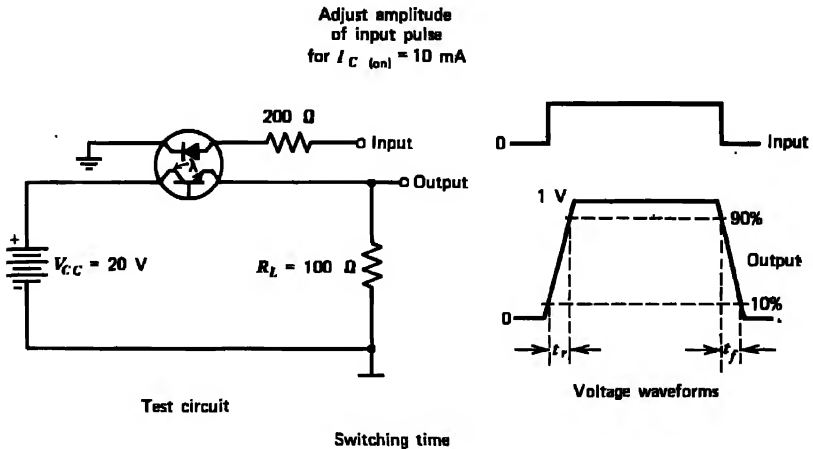


Figure 19. Optically coupled isolator. The input waveform is supplied by a generator with the following characteristics: $Z_{out} = 50 \Omega$, $t_r = \leq 100 \text{ nsec}$; $t_f = 50 \mu\text{sec}$, duty cycle $\approx 50\%$. The output waveform is monitored on an oscilloscope with the following characteristics: $t_r \leq 12 \text{ nsec}$, $R_{in} \geq 1 \text{ M}\Omega$, $C_{in} \leq 20 \text{ pF}$. (Courtesy of Texas Instruments, Inc.)

the optoelectronic pulse amplifier. This is an optically coupled pulse amplifier, in an integrated circuit package, consisting of a gallium arsenide light-emitting diode optically coupled to an integrated silicon photodetector feedback amplifier. The high input-output isolation of the optical coupling allows the device to function as a broad-band pulse transformer; its response extends down to zero frequency and it is compatible with other integrated circuits. Applications include transmission of AC or DC signals across computer or other subsystems where spurious currents prevent interconnecting of grounds and rejection of common-mode noise at the end of a long data-transmission line. Approximate size is $\frac{1}{4} \times \frac{3}{8} \times \frac{1}{8}$ in. Power rating is nominally 0.050 W (Figure 20). The biggest advantage of the solid-state relay over the conventional optoelectronic device is the reliability data and the variety of available types.

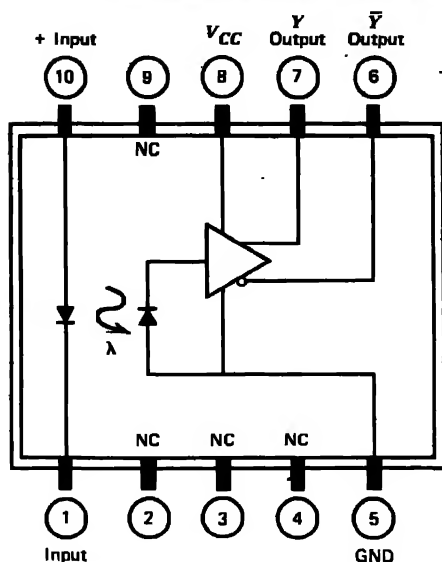


Figure 20. Optoelectronic pulse amplifier. Forward input polarity is indicated. NC = no internal connection. (Courtesy of Texas Instruments, Inc.)

5.9. STEPPER SWITCHES

Stepper switches are useful whenever a sequence of events is to be controlled or counted. They represent an intermediate stage between relay sequencing circuits and stepper motors. They are specifically designed for multistep programs that are difficult to synthesize with relays. Load ratings, high voltage capacity, and multipole capabilities are much greater than those available with solid-state circuits or stepper motors.

5.9.1. Construction

A stepper switch is composed of three principal parts: the electromagnetic circuit, the driving mechanism, and the contact assembly. The magnetic circuit functions in exactly the same way as a doorbell circuit. A battery, or source of DC power, is supplied to an electromagnetic coil. When the coil is energized, it attracts an armature. The armature in turn depresses a series of wafer-type switches or interrupter springs in series with the coil, thereby opening the circuit to the coil. When the coil is deenergized, the armature returns to its initial position, the interrupter springs reclose the circuit, and the process is repeated. The driving mechanism consists of a

ratchet, pawl, and driving spring that rotates the wiper assembly through a small increment each time the electromagnet is deenergized. It is important to note that the mechanism is cocked when the electromagnet is energized and it advances when the coil is deenergized. The brush assembly is composed of a rotating contact, or wiper, that is displaced by a fixed angular increment each time the magnet is deenergized, and a semicircular grouping of fixed contacts in a separate assembly (Figure 21). The fixed contacts are insulated from one another and also from the wiper. Each group of contacts and its associated wiper constitute a level and one discrete pole of the stepping switch. Most stepping switches contain more than one level; combinations of levels mechanically fastened together are called a bank. Levels may be composed of 13, 26, or 52 contacts, although there is no fixed standard on these numbers. Some applications require hundreds of contacts per level. The number of levels in a bank is also a variable, with up to 16 commonly used. The basic design of stepping switches requires the top level to be wired in a manner that will result in the wiper rotating through a specified interval; since the other wipers are ganged, or coupled to the top level, they will also rotate through the same number of contacts. In this way a number of independent poles are coupled to function together.

5.9.2. Circuitry

There are two general methods for using stepper switches: self-interrupted operation and impulse-controlled operation (Figures 22a and 22b). Self-interrupted operation using the doorbell circuit previously described, will cause the wiper to rotate from contact to contact so long as power is supplied to the magnet. Most units run at speeds up to 65 steps per second in this mode of operation. Impulse-controlled operation utilizes an external source of pulses to turn the magnet on and off. The interrupter spring is not used in this circuit. The pulses may be generated by a telephone dial or an electronic circuit. Speed is lower than self-interrupted circuits with a maximum speed of about 35 steps per second. Now that the methods of advancing the switch have been explained, the logical next question is how to stop the device? When external pulse sources are used, it is a simple matter to supply the correct number of pulses for a specified number of steps of the switch. It is not feasible to accurately control the number of pulses supplied or, if self-interrupted operation is used, the open or grounded potential contact technique may be employed (Figure 22c). The circuit shown is for self-interrupted switches, but a similar arrangement can be made for pulse-operated units. The coil is connected through the interrupter springs to one of the contacts on one level of the unit. All contacts are wired together, and also to a power source, except one called the "home" position.

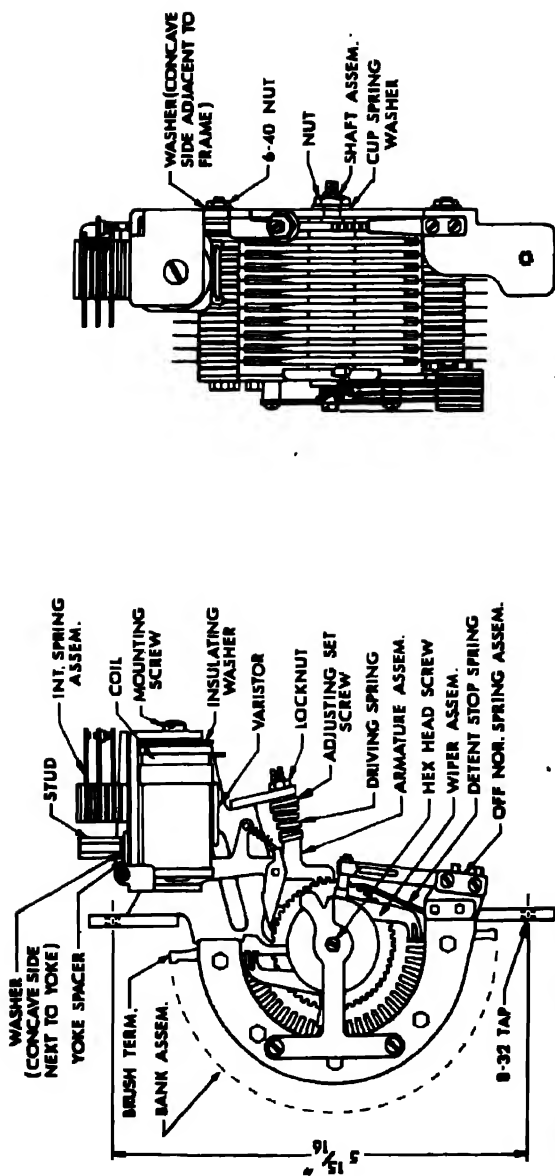
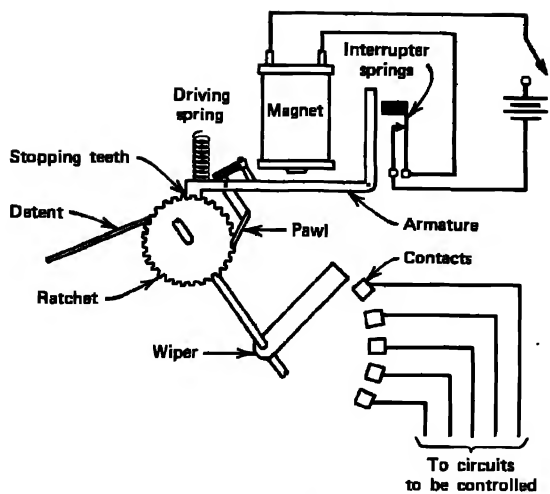
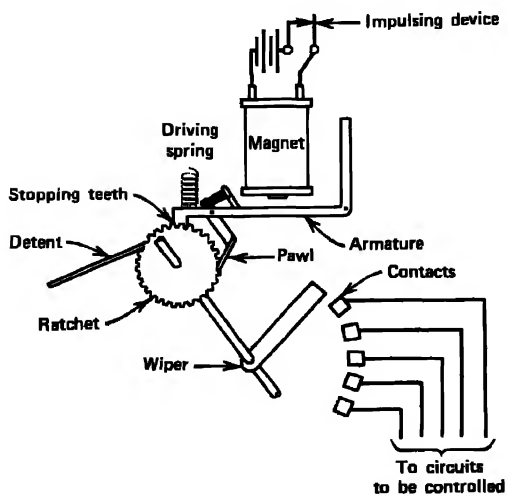


Figure 21. Stepper switch. (Courtesy of Automatic Electric Company. From Reference 9.)



(a)



(b)

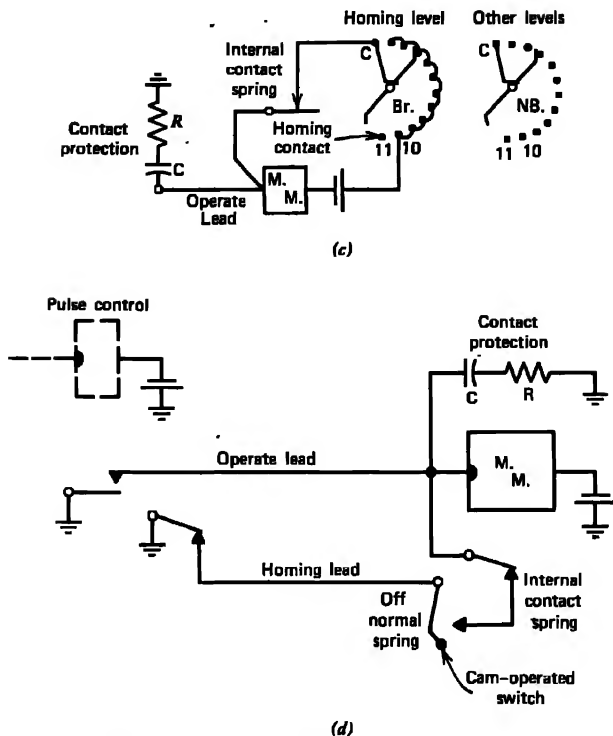


Figure 22. Impulse-controlled and self-interrupted stepper switches. M.M. signifies magnet motor or the relay coil. (a) Self-interrupted operation; (b) impulse-controlled operation; (c) homing through switch contacts; (d) homing through "off-normal" springs. (Courtesy of Automatic Electric Company. From Reference 9.)

When power is applied to the switch, it proceeds to move from one contact to the next until it reaches the home position where it stops because the point is at zero or ground potential. This is the most commonly used circuit in stepping switch applications because of the ease of changing the point where the unit will stop. The RC network shown is used for spark suppression only. It is also normal practice to use a varistor in parallel with a coil to decrease transients. All such circuits must be provided with some type of override or start switch to advance the switch off the home contact. Another method used to stop the switch at a precise point is the "off-normal" spring (or contact) method. This is simply an auxiliary switch in series with coil

that is actuated by a cam mounted on the wiper assembly. The cam can be rotated to the position corresponding to the desired stopping point (Figure 22*d*).

5.9.3. System Applications

Stepping switches have many functions, but they all fall logically into eight basic categories:

1. Control.
2. Time functions.
3. Counting.
4. Program.
5. Test.
6. Monitor.
7. Indicate.
8. Select.

The control function is the most important application in system design. Stepping switches can be directed by telephone dials, pulse generators, or any other voltage-shaping network. By using multilevel units almost any number of independent circuits can be controlled to function in synchronism as well as at fixed phase angles with respect to one another. The classic application is telephone switchboards where stepper switches have been developed to peak efficiency.

Time periods may be generated by using a homing contact circuit at a point that requires a fixed number of steps corresponding to the interval desired. The chief advantage of this system is that the interval can be changed quickly and precisely simply by grounding another contact. Most other methods require supplementary calibration to ensure the accuracy of the time period delivered.

Counting circuits function by using the last contact in a 10-contact level to supply a pulse to the coil of the next switch (Figure 23). This system pioneered the techniques later used in solid-state flip-flop circuits. The present use for these counting circuits is in application where the speed of solid-state circuitry is not required and the power levels are relatively high.

Program applications for stepping switches are numerous and interesting. The instrument used is equipped with a series of cams on each level that actuate "normal-off" contact (Figure 24). The circuit shown will function when the operate switch is closed and will drive cam 1 continuously. The "off-normal" switch will stop the action whenever the homing switch is closed and the cam is in position to engage it. Cam 2, in synchronism with cam 1, may be used to perform any desired control function. A typical

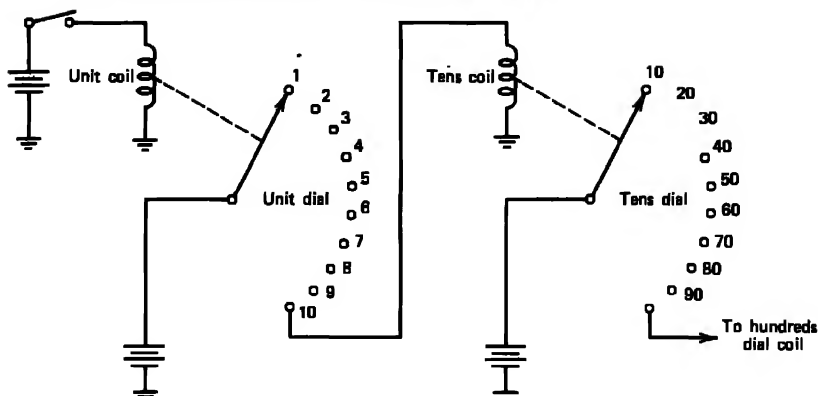


Figure 23. Counting circuit.

application is shown in Figure 25. The five cams have an output corresponding to 1, 2, 4, 8, 16, and 32 in binary code. Each cam performs a "divide by two" function. The decimal-to-binary conversion circuit shown in Figure 26 uses two separate, four-level stepping switches. Two channels of output are provided. The wiring on each level determines the output.

The testing function is illustrated by Figure 27. The circuit shows a method of automatically testing a six-conductor cable for continuity. If each conductor is continuous, the switch stops and proceeds to the next circuit; if it is open, the switch stops at that position. Most continuity checkers also have a second level connected in series with bulbs for each position to give the location of the open circuit. Practically, the cable is connected to the stepper switch by means of a "quick disconnect" connector. Typical applications involve 100 to 500 conductor cables that may be tested at rates up to 65 circuits per second. Numerous auxiliary recording devices also may be coupled to additional levels of the stepper switch. These include ammeters, ohmmeters, and similar instruments that are tied in with strip chart recorders.

Stepper switches also may be used to monitor a large number of circuits. The fixed contacts are connected to external circuits, and the wiper is wired to some form of recording or supervising device. When the measured variables exceed the prescribed tolerances, an auxiliary warning device on the recorder is actuated (Figure 28).

The indicator application consists simply of wiring a bulb to each contact position in a given level so that the progress of the stepper wiper can be observed easily. This method also can be used to actuate warning bells or control elements.



Figure 24a. Program-type stepping switch. (Courtesy of Automatic Electric Company. From Reference 9.)

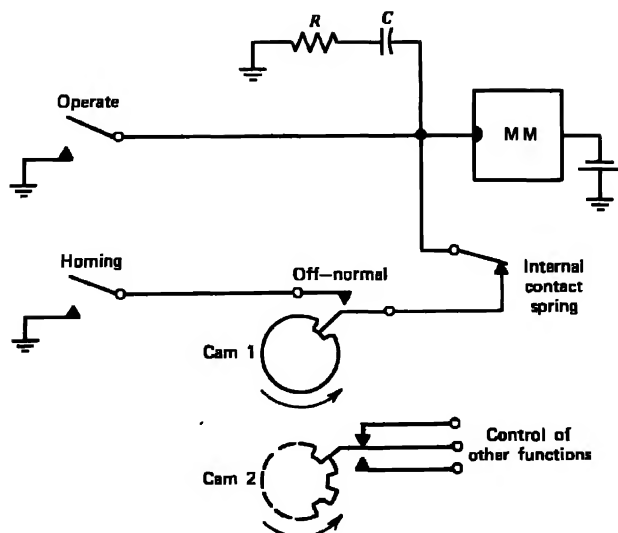


Figure 24b. Program switch operation. (Courtesy of Automatic Electric Company. From Reference 9.)

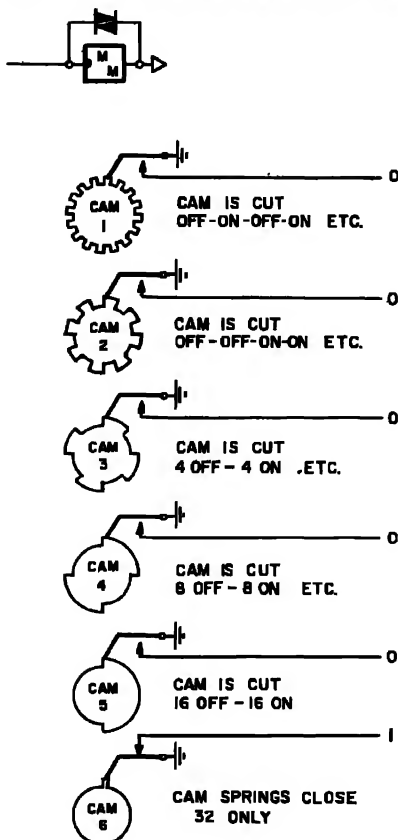


Figure 25. Program stepper switch used as a binary readout device. (Courtesy of Automatic Electric Company. From Reference 9.)

The last category, selection, consists of using an external pulsing device to advance the stepper switch to some desired position. This is particularly useful in experimental work where a complex circuit must be analyzed and certain junctions must be tested many times in random order.

5.9.4. Design Summary

Here are some of the requirements and pitfalls to look for when selecting a stepper switch.

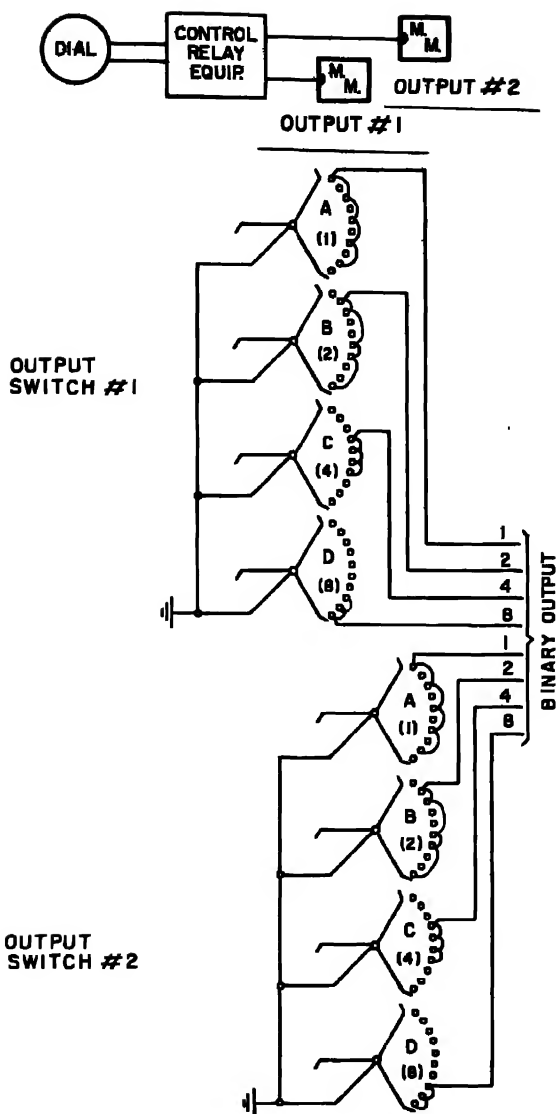


Figure 26. Decimal-to-binary conversion using stepper switches. (Courtesy of Automatic Electric Company. From Reference 9.)

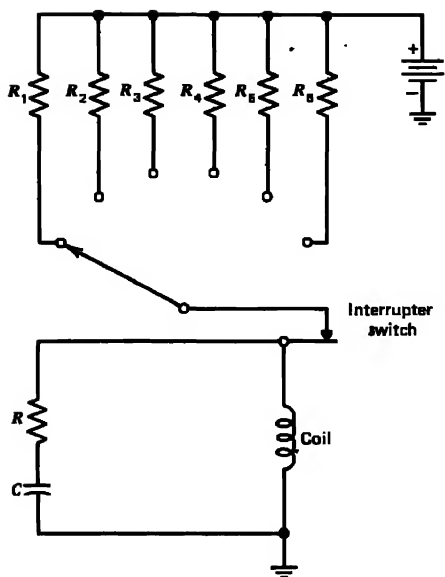


Figure 27. Continuity checker circuit.

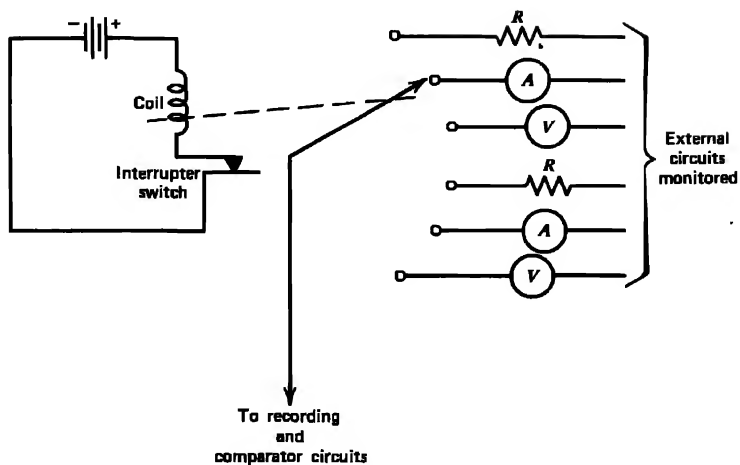


Figure 28. Multipurpose circuit checker.

The power supply used must be within $\pm 10\%$ of nominal rated voltage. Standard units are available in 6, 12, 24, 48, and 110 V DC. The use of rectified AC power is satisfactory, but operation will not be as smooth as with DC.

Coil ratings are for continuously pulsed operation. When the coil is to be held energized for long periods of time, a protective resistor must be inserted in the circuit. Coil ratings are limited normally to 500 V RMS. Wipers will withstand 1250 V RMS at most.

Contact ratings vary considerably, but a good value for minimum cost units is about 0.1 A at 115 V DC. If contacts are not used to interrupt current, ratings as high as 3 A may be employed. Contacts should always be protected with arc suppression networks. Varistors or RC networks are commonly utilized. Most contacts normally are open, but normally closed contacts are also available. They consist of two adjacent levels of contacts tensioned together so that each set of contacts forms a closed pair. The wiper assembly passes between them and opens them one at a time as it rotates.

Bridging or nonbridging wipers can be obtained. Bridging contacts "make before break" and nonbridging units "break before make."

Stepping switches are not designed to function at a precise speed. Each is adjusted by the manufacturer to work at an optimum speed consistent with such factors as coil inductance, stroke, spring tension, and magnetic characteristics. Two units from the same box may run at slightly different speeds. Although these devices are useful for generating timing pulses, they are by no means to be considered an electromechanical equivalent of an oscillator. The speed is also affected by temperature. Do not expect the speed to remain constant over a wide temperature range.

When stepper switches are to be operated in the self-interrupted mode, bridging-type wipers should be used to prevent burning the tips when breaking contact. For additional information on application pitfalls, the Automatic Electric Company book on *How to Use Rotary Stepping Switches* is highly recommended.

5.10. STEPPER MOTORS

The last part of this trilogy, devoted to stepper motors, represents a third stage in control techniques rather than an advance in relay or stepper switch design. The techniques are fundamentally different. Relays and stepper switches depend on programmed switch closures to control a process; stepper motors rely on shaft position for control functions. The basic bridge between the technologies is that they both utilize open-loop servotechniques for precise control.

A stepper motor is an electromechanical rotating device which, when energized by DC pulses in a programmed manner, indexes in precise angular increments. Its speed is proportional to the rate at which it is pulsed and can be determined as follows:

$$S_{\text{average}} = \frac{60(\text{pps})}{n}$$

where S_{average} = average shaft speed (revolutions per minute)

pps = pulses per second

n = number of phases in the winding

5.10.1. Types of Stepper Motors

Two basic types of stepper motors are in use--variable-reluctance and permanent-magnet types. Both units have stators with a number of wire-wound poles. The rotor on a variable-reluctance machine is composed of a number of lands and grooves, or slots and teeth. When the coils are energized, the teeth on the rotor align themselves with the poles on the stator (Figure 29). If the excitation is switched from one pole to another,

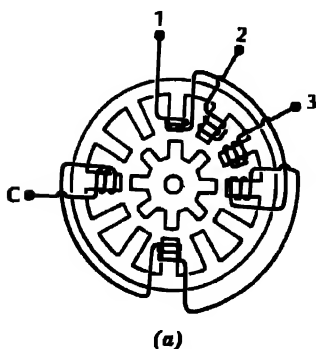


Figure 29a. Three-phase, 15° step, variable-reluctance motor construction. Complete winding shown on one phase only. (Courtesy of IMC Corporation. From Reference 11.)

the tooth on the stator closest to the second pole aligns itself there and turns the rotor a given angular increment. Therefore, nonadjacent poles must be pulsed successively in order to keep the rotor turning. This task is performed by logic elements in the driving circuitry.

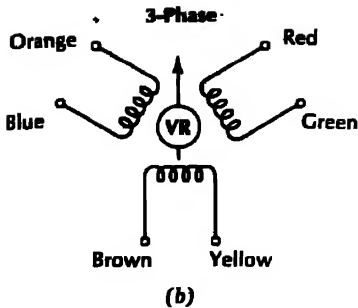


Figure 29b. Schematic of windings on a 15° variable-reluctance, three-phase stepper motor. (Courtesy of IMC Corporation. From Reference 11.)

The stepping motion of these devices is typical of an undamped second-order system. The initial displacement of the rotor overshoots the final position and gradually settles out (Figure 30). These oscillations are damped by the permanent magnet in the PM motor or by eddy currents in the VR motor. As the frequency of pulsing is increased, the period τ becomes shorter. When the period of the pulses is on the order of the oscillatory period t_1 , the unit has reached its maximum pulse rate.

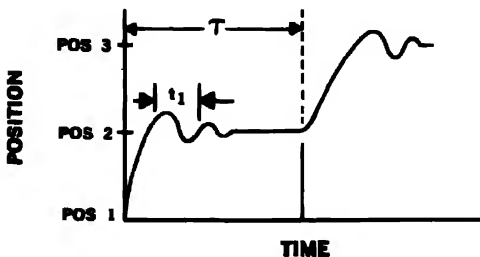


Figure 30. Stepper motor characteristic motion. The speed with which the unit moves is determined by $T = I\alpha$, where $T = T_{\max} \sin \psi$, I = total inertia, and α = angular acceleration. ψ = electrical angle of rotation. (Courtesy of IMC Corporation. From Reference 11.)

The operation of a VR motor is shown schematically in Figure 31. When winding *C* is energized with DC, the rotor aligns itself with the resultant pole *D*. When winding *B* is energized, the closest pole *G* moves 30° into alignment. When winding *A* is energized, pole *F*, 30° away, moves into alignment; and so on. Under these conditions stepping angles of 30°

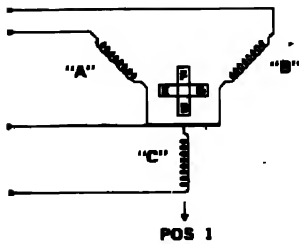


Figure 31. Operation of a variable-reluctance stepper motor. (Courtesy of IMC Corporation. From Reference 11.)

are obtained. The polarity is not important, since the rotor moves to a minimum reluctance position. The winding excitation for a three-phase unit is shown in Figure 32, and the winding excitation for a two-phase appears in Figure 33.

Figure 34 illustrates two ways of obtaining 15 steps. The first method uses a four-pole stator and an eight-pole rotor and results in motion to the left.

Excitation Mode	Energized Winding	Rotor Position	Motion Sequence
3-Phase commutation of B+ only	2-1	a	CCW
	3-4	c	CCW
	5-6	e	CCW
3-Phase modified commutation of B+ only	2-1 & 3-4	b	CCW
	3-4 & 5-6	d	CCW
	5-6 & 2-1	f	CCW

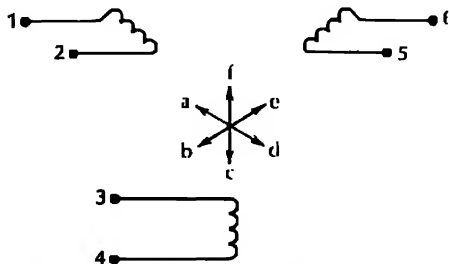


Figure 32. Three-phase and modified three-phase excitation. The modified three-phase mode simultaneously excites two phases and the rotor indexes to a minimum reluctance position corresponding to the resultant of the two magnetic fields. This improves damping characteristics but increases power dissipation. (Courtesy of IMC Corporation. From Reference 11.)

Excitation Mode	Energized Winding	Rotor Position	Motion Sequence
2-Phase commutation of B+ and B—	3-1	f	Index
	6-4	h	CCW
	1-3	b	CCW
	4-6	d	CCW
2-Phase modified commutation of B+ and B—	3-1 & 6-4	g	Index
	1-3 & 6-4	a	CCW
	1-3 & 4-6	c	CCW
	3-1 & 4-6	e	CCW

1



3

Figure 33. Two-phase and modified two-phase excitation. One entire phase of the motor, end-tap to end-tap, is energized at a given moment in time; the modified two-phase mode simultaneously excites both windings (considering one winding to be end-tap to end-tap and ignoring the center taps). If the center taps are used, the unit can be run as a four-phase machine. (Courtesy of IMC Corporation. From Reference 11.)

The second method utilizes an eight-pole stator and an eight-pole rotor and moves to the right. The first method is preferred, since it results in fewer slots and coils and is less expensive to build.

The permanent-magnet rotor is a cylindrical permanent magnet. Exciting a stator winding results in the formation of two magnetic stator poles, one North and one South. The rotor aligns itself with this magnetic field. In the two-phase winding where two windings are excited simultaneously, two poles are developed by each phase, but added vectorially, only one North-South field of two poles exists. In this case, the rotor aligns its poles with the resultant field halfway between the excited winding (Figure 35). The use of a two-pole-per-phase stator winding and a two-pole rotor magnet results in four discrete steps per revolution; the rotor follows each 90° shift of the stator field (Figure 36). To obtain a 45-step angle, a four-pole stator and a four-pole rotor magnet must be used (Figure 37). Therefore the number of poles in the stator and the rotor can be varied to produce the desired number

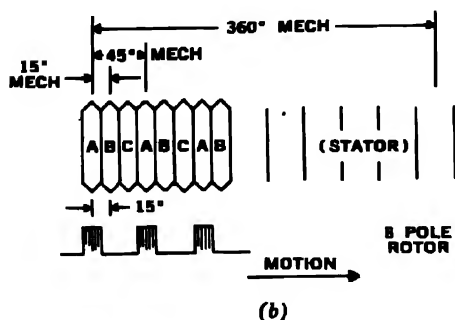
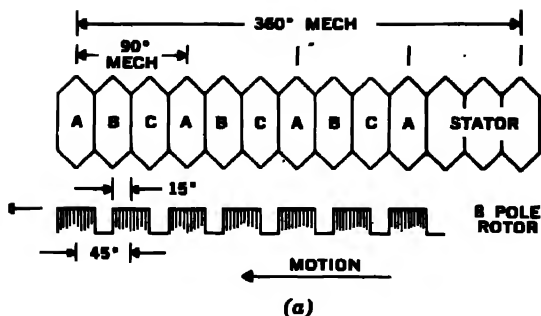


Figure 34. Two methods of obtaining 15° step increments from a variable-reluctance motor. (a) Four-pole rotor and 8-pole rotor. (b) Eight-pole rotor and eight-pole rotor. (Courtesy of ICM Corporation. From Reference 11.)

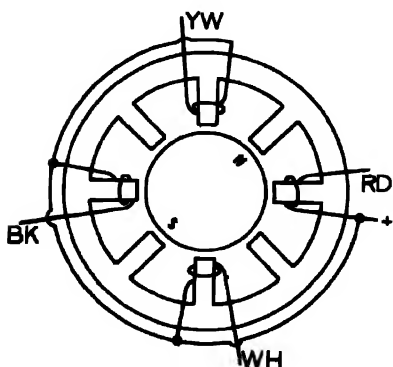


Figure 35. Two-phase permanent magnet motor. (Courtesy of Singer-General Precision Inc. From Reference 10.)

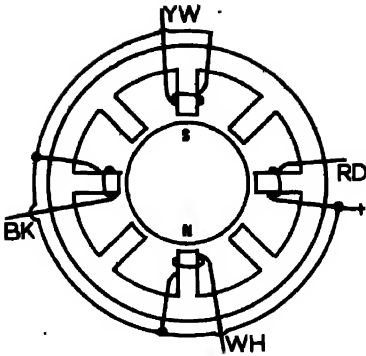


Figure 36. Four-phase permanent magnet motor. (Courtesy of Singer-General Precision, Inc. From Reference 10.)

of steps per revolution. Machine design practice limits the number of steps to eight per revolution for a PM machine.

Consider the methods of excitation outlined in Figure 38. There are four different methods of energizing a two-pole, four-phase PM stepper motor. Three of the methods produce 4 steps per revolution and one method produces 8 steps per revolution. Some of the sequences may appear to be equivalent, but their efficiencies may vary by a factor of 4-1.

The relationship between torque and pulse rate is shown in Figure 39. In general, a system should not be designed that requires more than half the stall torque. The exact shape of the torque curve is a function of the number of poles, phases, and other factors and should be obtained from the manufacturer for each unit.

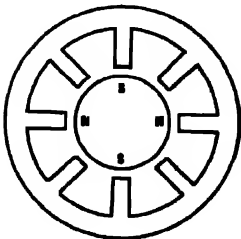
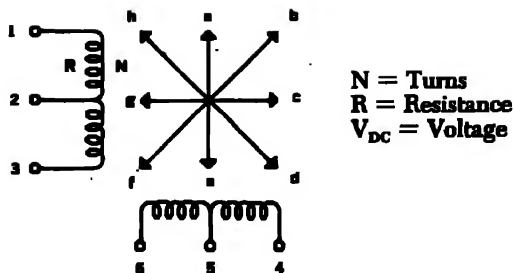


Figure 37. Four-pole stator and four-pole rotor permanent magnet motor. (Courtesy of Singer-General Precision, Inc. From Reference 10.)



	WINDINGS TO WHICH VOLTAGE IS APPLIED	POSITION	MOTION
Excitation- Technique A	V_{1-2} and V_{4-5}	b	—
	V_{3-2} and V_{4-5}	d	90° CW
	V_{3-2} and V_{6-5}	f	90° CW
	V_{1-2} and V_{4-5}	h	90° CW
Excitation- Technique B	V_{1-3}	a	—
	V_{4-6}	c	90° CW
	V_{3-1}	e	90° CW
	V_{6-4}	g	90° CW
Excitation- Technique C	V_{1-2}	a	—
	V_{1-2} and V_{4-5}	b	45° CW
	V_{4-5}	c	45° CW
	V_{4-5} and V_{3-2}	d	45° CW
	V_{3-2}	e	45° CW
Energizing one winding then 2 windings in parallel and then 1 winding, etc.	$V_{(1-4)}$ (3,6 CONNECTED)	h	—
Excitation- Technique D	$V_{(1-6)}$ (3,4 CONNECTED)	b	90° CW
	$V_{(4-1)}$ (3,6 CONNECTED)	d	90° CW
	$V_{(6-1)}$ (3,4 CONNECTED)	f	90° CW

Figure 38. Possible methods of exciting a permanent magnet stepper motor. (Courtesy of IMC Corporation. From Reference 11.)

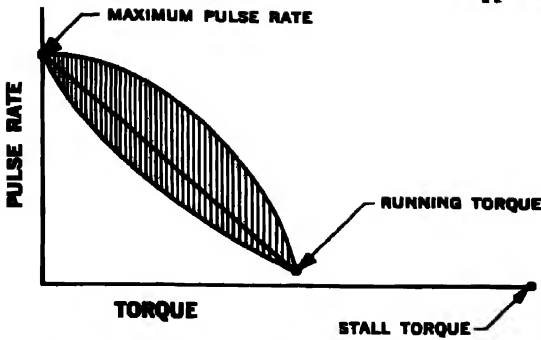


Figure 39. Pulse rate versus torque for stepper motors. (Courtesy of IMC Corporation. From Reference 11.)

The relationship between the inertia of the load and motor inertia is shown in Figure 40.

In summary, PM units should be selected when:

1. Large stepping angles are desired.
2. Pulse rate is low (300 pps maximum).
3. Magnetic detenting is required.

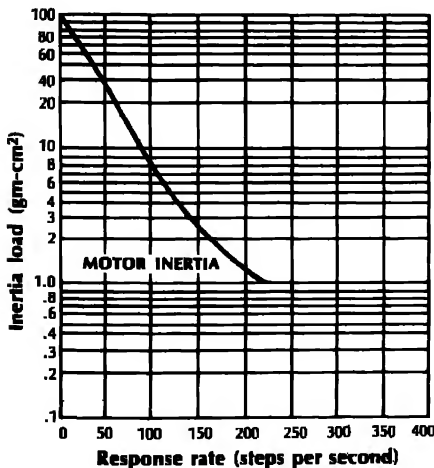


Figure 40. Inertia load versus stepping rate for 90° permanent magnet stepper motor. (Courtesy of IMC Corporation. From Reference 11.)

Variable reluctance units should be used when:

1. Pulse rate is high (1200 pps maximum).
2. Small angular steps are required.
3. Magnet detenting is not required.

The principal problem associated with both types of stepper motors is the tendency to overshoot the position, which can be very damaging to precision control systems. This is their most serious limitation when compared with precision servosystems.

5.10.2. Drive Circuits

The function of the drive circuit is to energize the windings of the stepper motor in the proper sequence. For each combination of poles and slots there is an optimum order for energizing them. Once the program has been determined, the circuit must conform to it without variation. The input to the drive circuit is a train of pulses generated by some form of fixed-frequency pulse generator. Typical circuitry is shown by Figure 41. Circuit components are composed typically of solid-state or integrated components. The circuit shown in Figure 41a is designed to energize windings in a 1-2-3-4 sequence. The motor may be operated in the CW or CCW direction by changing the selector switch. Each power transistor is controlled by an "and" circuit. Two flip-flop units generate the pulses to the four gates. The control logic is shown below the circuitry. The logic for a three-phase unit is shown in Figure 41b. A constant current power supply for stepper motors is highly recommended.

5.10.3. Gearing

Many applications require any odd number of steps per revolution or torque characteristics that are not available in off-the-shelf items. Some other reasons for using gearing are:

1. To change the output speed.
2. To match the inertia of the load and the motor.
3. To convert the rotary motion to the linear motion through a gear and rack arrangement.

The design of gearing to be used with stepper motors should be on the conservative side. Ordinary motors turn in a continuous manner and the

loading is even. Stepper motors turn in a series of steps; consequently, the gearing is subjected to a series of shock loads. It is a good idea to invest in more precise gearing since the small increments available with variable reluctance units may be partially compromised in the errors inherent in the cheaper gear trains. This generally can be accomplished by buying gears one commercial grade higher than normal. The extra cost usually is justified.

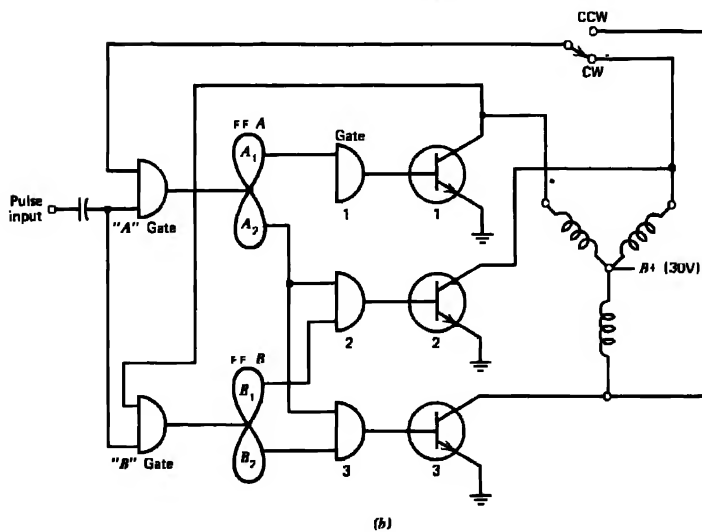
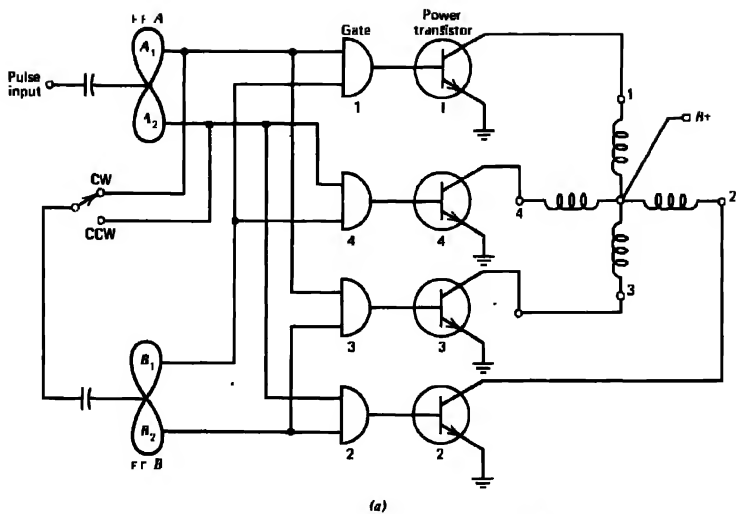
5.10.4. Applications

The most important use of stepper motors is in open-loop control systems. The basic formula is "one pulse, one motor step." This approach eliminates the costly feedback components that are standard with conventional servo-systems. It eliminates feedback transducers, amplifiers, integrators, A/D converters, and much associated engineering analysis. The decreased number of components also contributes to lower costs and improved reliability.

The first criterion in determining if an open-loop system is acceptable is the allowable system error. The resolution of an open-loop stepper motor system is the number of steps per revolution or the number of degrees per step. This must be consistent with the overall system requirement before open-loop control can be considered. Next, the tolerance on this number must be taken into account; it is normally ± 3 to 5% of the step increment but is noncumulative. This means that any *one* step may be as much as 5% in error; successive steps do not necessarily have this error. Like any other motor, the speed at no-load and full load are different; therefore, the number of pulses per second will vary under load. This should not be considered an error but rather an engineering problem that must be recognized. (Refer back to Figures 39 and 40.)

The second basic criterion in determining if open-loop control is feasible is the inertial load. The system must be designed so that the motor can be started and stopped within the allowable system errors. All stepping motors have a range of speeds and loads within which they will start and stop without losing count of pulses. If the inertial load is not in this range, consider gearing.

The third criterion involves analyzing what might happen to the system without feedback functions if unexpected inputs, such as extreme temperature, power failure, and shock, occurred. Would an uncontrolled system lead to catastrophic results? Many systems can be controlled by simple limit switches, rather than servos, to avoid serious problems. Discrete control systems will serve this function. The last criterion (but usually considered first) is economic. Would the system be too costly if built with conventional



servos? Perhaps a less accurate open-loop approach would make the system salable. (Figure 42.)

The discrete feedback control system is an economic compromise between the open-loop and closed-loop systems. It contains a feedback element that provides "yes-no" type of information; this indicates that the system has changed its state and corrective action must be taken. Examples are position switches, pressure switches, and temperature switches. They are all intermittent in nature while conventional feedback devices are continuously proportional to some analog or digital process. Elevators and conveyors are examples of systems that profit by utilizing discrete feedback systems. Another advantageous technique in discrete systems is "slewing." Slewing can be defined as motion so much faster than the normal rate that standard control functions do not apply (for example, quick return mechanisms). The unit cannot stop, start, or reverse on a given pulse in this mode of operation.

5.10.5. Closed-Loop Applications

Although the main usage of stepper motors today is in open-loop systems, they are by no means excluded from feedback servosystems. On the contrary, stepper motors have many inherent advantages over older types of motors. The dynamic starting and stopping characteristics of this device make it almost ideal for high-performance systems, since they do not hunt or oscillate. The most sophisticated systems pair incremental encoders and stepper motors as an all-digital sensor and driver package. The advantages

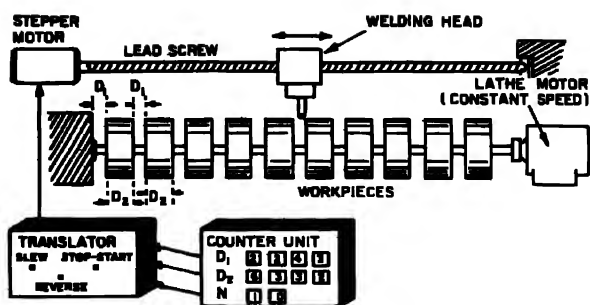
Figure 41. Typical logic circuitry used with stepper motors. (a) Four-phase logic. Control of the four stages (hence the four motor windings) is determined by the states of flip-flops *A* and *B* as follows:

Flip-flop state	Power stage condition
<i>AB</i>	Transistor 1 is "on"
$\overline{A}B$	Transistor 2 is "on"
<i>A}\overline{B}</i>	Transistor 3 is "on"
$\overline{A}\overline{B}$	Transistor 4 is "on"

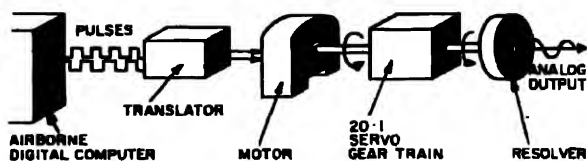
(b) Three-phase logic. Control of the three power stages (hence the three motor windings) is determined by the state of the power transistors in the following manner:

Flip-flop state	Power stage condition
<i>A</i>	Transistor 1 is "on"
$\overline{A}B$	Transistor 2 is "on"
$\overline{A}\overline{B}$	Transistor 3 is "on"

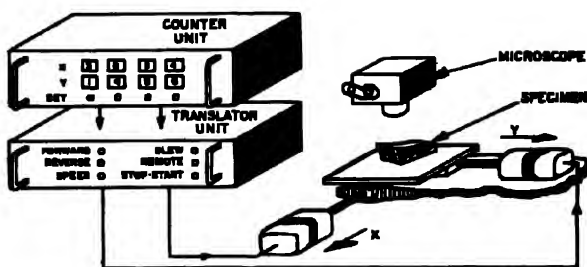
(Courtesy of Benwill Publishing Company. From Reference 20.)



(a)



(b)



(c)

Figure 42. Open-loop applications of stepper motors.

(a) Welding lathe positioner. *Problem.* Position a precision welding head over a welding lathe bed for the fabrication of cylindrical pressure vessels. System must have two programmable distances D_1 and D_2 , so that the work cycle can be performed as follows: start; traverse D_1 ; perform weld; traverse D_2 ; perform weld; traverse D_1 ; and so on. This cycle is repeated a preset number of times and then stops; the welding head returns automatically to its starting position. The machine is unloaded, reloaded with fresh material, and restarted by an attendant who operates four such machines. *Application.* This system was open-loop controlled by:

Stepping motor	Superior Electric	Model SS-100
Translator	Icon Corporation	Model 440
Counter	Post Electronics	Special

The motor was selected for its torque output at start and stop; the translator for its high driving power, panel-or-remote input provisions and a fast reverse slew mode; the counter for packaging flexibility.

(b) Digital-to-analog converter. *Problem.* Convert computer angles from a digital airborne computer to analog form, in terms of a continuous cosine function. Power is not a factor; speed, accuracy, and economy are important. *Application.* This system was open-loop controlled by:

Stepping motor	Sigma Cyclonome	Model 9AB2
Translator	Sigma Cyclopulser	Model 9FO1
Resolver	Standard servo grade	
Geartrain	Standard servo grade	

The motor and translator were selected for their relative economy.

(c) Metallurgical specimen positioner. *Problem.* Position a metallurgical specimen beneath a crystallographic microscope of a very limited field in such a way that you can find your way back to a selected spot without scanning the entire specimen again. The system must move 2 in. in both X and Y directions, have a resolution of ± 0.1 mil, and slew at $\frac{1}{4}$ in./sec. It must have visual coordinate readout and the ability to scan a preset square grid. Both axes require manual override. *Application.* This system was controlled by:

Two motors	United Shoe Machinery Corporation	Model HDUM-15-100
Translator	Icon Corporation	Model 420
Counter	Modular Instruments Counting Modules	

The motors were selected for their high resolution (800 steps/rev as compared with the usual range of 4 to 200 steps/rev) and their high holding torque for maintaining X -position while running in the Y -direction. The translator was selected for its power and because its built-in logic enables it to "remember" direction as well as to drive both motors from a common power supply. The need for modular flexibility determined the counter choice. It was necessary to control the two axes with three 5-digit decade counters, since the scan grid proceeds in N "swaths" of width X and Y , all of which are variable. Position readouts were necessary for only two numbers (X and Y). (Courtesy of Icon Corporation. From Reference 12.)

of this approach are good immunity to noise, no servoamplifiers, and excellent coupling between the sensor and motor.

The conventional use of stepper motors in closed-loop servosystems is illustrated in Figure 43. It features components that have been thoroughly debugged and have a long history of successful applications. This system is significantly costlier than open-loop systems, but it is inherently more accurate and versatile. Compared to digital-type servosystems, it is cheaper and its technology is better developed (in 1971).

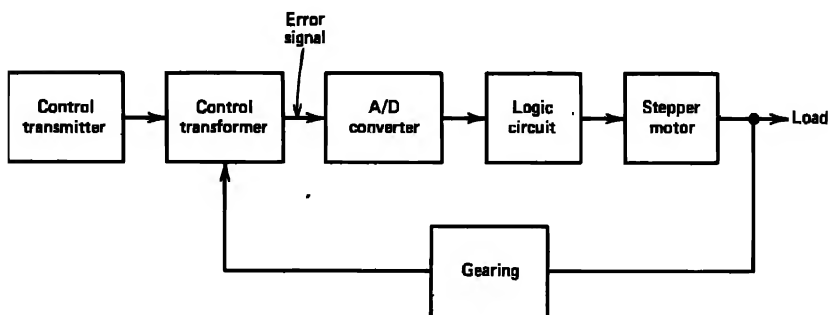


Figure 43. Closed-loop stepper motor servo.

5.10.6. Summary

In a discussion of this type, some technical details and definitions inadvertently are "swept under the rug," because they do not belong to any particular topic. Nevertheless, they should be included in the discussion. A few of them are examined below.

MAGNETIC DETENT. Permanent-magnet stepper motors have a rotor that is permanently magnetized; consequently, after the stator windings are deenergized, the rotor is still attracted to the pole opposite it on the stator. The torque is substantial. Rated torque is the increment above detent torque that is available for accelerating the rotor when the coils are energized. Detent torque is useful in eliminating the ambiguity of rotor position at startup. Variable reluctance units may or may not remain at their last position after deenergization occurs, since any shaking force acting on an unbalanced rotor will cause it to shift. Detent torque is sometimes referred to as a "memory" characteristic; it is also called holding torque.

ROTATION. Stepper motors can be obtained that operate in one direction only or in either direction. Unidirectional units are less expensive, since the

driver circuit has fewer components. The motor is essentially the same in both cases.

PULSE STEPPING. Pulse stepping is a technique used on single-phase units where one pulse causes the unit to rotate 360° . It eliminates logic circuit completely. The instrument generally is equipped with gearing so that the output of the shaft can be almost any angular increment desired. Output rates up to 60 pps are commercially available. One design, marketed by the Hayden Switch and Instrument Company, indexes 180° when power is applied and 180° when power is removed, thereby providing finer resolution.

MAXIMUM RESPONSE. The maximum pulse rate that can be applied to a step-servo in a random manner (CW and CCW), resulting in synchronized steps, is maximum response.

SLEW RANGE. This is the high-speed range in which the motor can run continuously but cannot stop, start, or reverse without losing step count. It is used when a fast return or approximate positioning is required. The speed must be reduced below the slew range before normal control is again possible.

POWER ANGLE. The angle of lag between the rotor and the axis of the accelerating magnetic field is the power angle. It is usually specified under full load.

PULSE RATE. Pulse rate is the number of pulses per second (pps) applied to the starting winding. Maximum pulse rate is the fastest pulse rate at which an output torque can be obtained (Figure 39).

MAXIMUM RUNNING TORQUE. This is the maximum torque that can be delivered at the shaft when the motor windings are sequentially energized.

STALL TORQUE. The maximum torque that a unit can deliver when the motor windings are energized at zero pulse rate is called stall torque. It cannot be delivered under dynamic conditions.

5.10.7. Closure

Although the stepper motor seems to be an almost perfect tool, there remains one "Achilles' heel"—mechanical resonance. Mechanical resonance usually occurs between the upper synchronized speed range and the slow range. The symptoms are sudden nonuniform rotation and, sometimes, reversal of rotation. Above and below the critical speed no problem is apparent. The critical speed is basically a function of the motor-load elasticity and damping. No complete analysis of this problem is presently

available, but it is being studied by virtually all stepper motor manufacturers. See References 16 and 17. Fortunately, several methods have been developed for keeping the problem under control. These include mechanical, electrical, and viscous damping techniques. Each company has its preferred techniques, but to date no universally accepted "fix" has been agreed upon. This problem has not been widely advertised. A more detailed discussion of the problem can be found in references given below.

REFERENCES

1. *Definition of Relay Terms*, National Association of Relay Manufacturers (NARM), Scottsdale, Ariz., 1964.
2. *Engineers Relay Handbook*, NARM, Scottsdale, Ariz., 1964.
3. *Seventh Annual Relay Conference Proceedings*, NARM, Scottsdale, Ariz., 1964.
4. H. C. Roters, *Electromagnetic devices*, Chapter 13, John Wiley & Sons, Inc., New York, 1941.
5. Relay Applications, *Electronic Design*, November 29, 1965, Hayden Book Company, New York.
6. *Basic Relay Circuits*, Automatic Electric Company, Northlake, Ill., 1965.
7. *Relay Magic*, Automatic Electric Company, Northlake, Ill., 1965.
8. *Reed Relays*, Wheelock Signals, Inc., Long Branch, N.J., 1966.
9. *How to Use Stepping Switches*, Automatic Electric Company, Northlake, Ill., 1965.
10. *Technical Information for the Engineer*, Singer—General Precision, Inc., Little Falls, N.J., 1968.
11. *Tormax Steppers and Controls*, IMC Magnetics Corporation, Maywood, Calif., 1970.
12. *Guidelines for Specifying Stepping Positioning Systems*, Icon Corporation, Cambridge, Mass., 1968.
13. B. F. Papa, Analyzing Sampled Data Servos Using Stepper Motors, *Electrotechnology*, February 1969, New York.
14. T. R. Fredriksen, Closed Loop Stepping Motors, *Machine Design Magazine*, September 16, 1965, Penton Publishing Company, Cleveland, Ohio.
15. *Slo-Syn Stepping Motors*, Superior Electric Company, Bristol, Conn., 1970.
16. A. S. Audeh and A. N. Deodhar, Speed Control: Four Methods Compared, *Control Engineering*, December 1970, Reuben H. Donnelley Company, New York.
17. H. Waagen, Reduce Torsional Resonance in Incremental Servos, part of reprint #944, Stepping Motors and Incremental Servos, *Control Engineering*, Reuben H. Donnelley Company, New York, 1970.
18. *Electromechanical Design*, April 1959, Benwill Publishing Company, Brookline, Mass.
19. *Machine Design Magazine*, September 30, 1965, Penton Publishing Company, Cleveland, Ohio.
20. *Electromechanical Design*, July 1965, Benwill Publishing Company, Brookline, Mass.

Chapter VI Slip Rings and Potentiometers

Slip rings and potentiometers are critical system components because they control the power input and scale factors of the system. Ideally, they should be completely passive devices. In practice, each component consumes power, generates noise and heat, and may introduce discontinuities and scale factor errors in the system. The problem common to both components is efficient commutation. The proper selection of available hardware is difficult without a detailed knowledge of the metallurgical, mechanical, and electrical solutions to this problem.

6.1. DESIGN OBJECTIVES

The design objectives for good commutation are low contact resistance, prevention of erosion and metallic transfer from one surface to another, and elimination of welding tendencies.

The contact resistance across a closed pair of contacting surfaces is a function of the contact geometry, a constriction resistance, and any intermediate film resistance. Two surfaces that touch are usually designed for mutual tangency to eliminate voids between them. The most efficient geometry for contacting surfaces is two radii, or at least one radius and a smooth surface. Constriction resistance is the condition where only a small part of the contacting surfaces actually touch each other, thereby narrowing the effective cross-sectional area; the resistance is higher than nominal. Film resistance is a function of impurities deposited on the mating contact surfaces. It is caused by condensation of airborne impurities on the critical surfaces. Outgassing of components in confined envelopes is another source of contaminants. The only way of guarding against contaminants is to scrupulously clean the parts at assembly and keep them in a hermetically sealed package. Increasing the contact pressure is also beneficial, but this

accelerates wear. The power loss associated with voltage drop across contacts can be a significant source of parasitic heat. Variations in contact resistance due to varying pressure and changing film conditions can introduce a degree of instability in a normally "well behaved" servosystem. The following equation sums up the situation:*

$$R_c = K\rho\sqrt{H/F} \quad (1)$$

where R_c = contact resistance

K = design constant

ρ = specific resistance of the contact material

H = contact hardness

F = applied contact load

Electrical erosion is caused by arcing across the contacts. The heat generated by the arc is sufficient to bring the contacts to the boiling point; metal is then lost by vaporization. The higher the current, the greater the vaporization loss. When designing a circuit that uses some form of DC commutation, it is desirable to extinguish or open the contact as quickly as possible to minimize the arcing problem. AC circuits are self-extinguishing, since the voltage waveform passes through zero potential in every cycle. The metallurgical approach to this problem is to select a contact material with a high boiling point, high specific heat, and high thermal conductivity. Materials such as silver and silver cadmium oxide meet these requirements.

Transfer is the migration of contact material from one contact to another. It leaves a crater in one contact and forms a buildup on the opposing contact. DC circuits are the most troublesome. Transfer to the negative contact is called bridge transfer, and transfer to the positive contact is called arc transfer. Bridge transfer occurs when contacts operate under nonarcing conditions. When the contacts separate, the contact resistance rises rapidly because of the decreasing contact force; I^2R heating causes the last point of contact to be heated to the vaporization temperature. A molten bridge of metal forms between the contacts, and asymmetry in the heating and rupture process causes the metal to migrate and build up on the face of the negative contact. The rate of transfer is proportional to the current. When contacts arc on opening the circuit, due to high circuit voltages or induced voltages, there is a tendency to transfer metal to the positive contact. If this process is of greater influence than the negative transfer, the result is a buildup of metal on the positive contact. Rectifiers, capacitors, and resistors are used to minimize positive transfer. Negative

* Source: *Mallory Electrical Contacts Handbook*.

transfer is controlled by adding inductance in the circuit. The metallurgical approach is to pair dissimilar materials with high vaporization temperatures such as tungsten and platinum rhenium or silver cadmium oxide and silver alloys.

Welding or sticking of contact surfaces is caused by an arc melting metal on the contact surfaces. When the arc is extinguished, after the closing of contacts, the molten metal solidifies and welds the two surfaces together. This condition is common in motor-starting circuits and other applications with high starting currents. Contact bounce also intensifies this condition. Welding can be minimized by increasing contact pressure, thereby eliminating contact bounce and limiting inrush current. Preferred contact materials are those with good thermal conductivity, high melting point, and poor wettability such as silver cadmium oxide, silver iron, and silver graphite.

The three prime causes of commutation failure are excessive heat, mechanical wear, and environmental contamination. Heat may develop by excessive voltage drop across the contacts due to any of the four conditions just described. It is also caused by I^2R loss when the cross section of the contact assembly is too small. The type of load is also important. "On-off" inductive loads, such as motor-starting circuits, are much more demanding than purely resistive loads. Repetitive inductive loads are the most difficult ones to handle reliably. Arc suppression, circuits using rectifiers, and diodes are necessary to prevent excessive heat buildup in these types of applications. Some heat concentration is unavoidable in even the most ideal applications. The technique of conducting this heat away, or heat-sinking, is a prime design problem. If the coil is wound on a mandrel, as in potentiometers, the thermal-conductivity of the mandrel should be as high as possible. This leads to a problem, since good thermal conductors usually are good electrical conductors as well and would produce shorts between adjacent coils. Successful design is always a compromise between these two factors. Contact assemblies have a similar problem—the backing material must be capable of conducting away the heat generated and yet remain electrically insulated. This problem is discussed further in Section 6.3.2.

Mechanical wear is caused by unequal hardness, contact bounce, and contamination of the contact surfaces. Selection of materials with compatible hardness is the easiest part of the problem. Elimination of contact bounce is somewhat more difficult. The wiper or contacts are normally preloaded to minimize bounce but, under severe conditions of shock or vibration, some bounce will occur. To eliminate it completely requires increasing the contact pressure, which, in turn, causes more rapid wear of parts and seriously limits the life of the device. This is another design compromise situation. When dirt or other contaminants find their way between contacting surfaces, not only is the electrical path interrupted, but the contacts are spread

farther apart, the preload and arcing are increased, and electrical noise is generated. Very often the contaminant is harder than the contact, the highly polished surface becomes irreparably scoured, and the driving torque for the device increases.

Environmental contamination includes the solid contaminants discussed above as well as the condensation of gaseous products on contacts and commutating surfaces. The film produced by condensation of gases results in an increase in contact resistance and, therefore, in excessive temperature gradients. These films may be generated by outgassing of components, oxidation, and sulfidation. Outgassing of components is common under high vacuums such as those encountered in aerospace applications. Oxidation is not a serious problem when metals such as gold, platinum, and palladium are part of the contact alloy. Unalloyed silver is subject to oxidation, which severely limits its usefulness. Hydrogen sulfide in the air is responsible for a large part of the sulfurous deposits found on contact surfaces. Carbonaceous deposits generated by insulation breakdown often form relatively thick deposits on unprotected surfaces. Severe cases lead to complete disruption of commutation.

6.2. METALLURGY

Selection of material for contacts used in slip rings and potentiometers follows similar guidelines:

1. High electric and thermal conductivity.
2. High oxidation resistance.
3. High melting point.
4. Good hardness.
5. Good tensile strength.

We discuss specific alloys after the design of each component has been examined.

6.3. SLIP RINGS

The best way to start a discussion of slip rings is to consider the system parameters that must be determined before selecting a particular unit. The key electrical parameters are:

1. Number of independent circuits and commons.

2. Current and voltage per circuit.
3. Contact resistance is substantially different for a slip ring in motion than for one at rest. These two parameters must be specified at definite current loads.
4. Leakage current is generally dictated by system specifications tied to one of the familiar government specs such as MIL-E-5272 or MIL-E-5400. Standard practice includes a megger check at 500 V DC and a 1000 V AC "hipot" test. Commercial applications generally follow the same procedures.
5. Electrical noise is given in terms of a peak-to-peak voltage at a specific load current and bandwidth. Noise usually decreases after an initial run-in time. Noise voltage is approximately proportional to the change in contact resistance for current values below 0.1 A. High-quality instrument slip rings exhibit noise values as low as 5 μ V. More moderately priced units are rated at 10 to 100 μ V.
6. Cross talk is a parasitic phenomenon defined as the ratio of the signal voltage in one circuit to the voltage induced in an adjacent circuit (in decibels). Cross talk is a function of conductor spacing, length, resistivity; the dielectric constant of the insulation and the proximity of adjacent rings are also prime factors. At high frequencies circuit-to-circuit capacitance is also important. Typical cross talk ratios are about 70 to 80 dB at frequencies up to about 100 kHz and 40 to 50 dB at 50 MHz; typical load for these values is about 200 Ω .

Some of the mechanical system parameters that must also be specified are the following:

7. Available envelope; as in all other instruments, cost is usually inversely proportional to the required size.
8. Shaft speed—the slower, the better, because at low speeds and low current, wipers can be made of relatively soft alloys that minimize noise. As the surface speed increases, harder materials are required. Most manufacturers group surface velocity in three categories: low speed—up to 250 ft/min; intermediate speed—250 to 5000 ft/min; high speed—above 5000 ft/min.
9. Torque is computed from the equation

$$T = NFR\mu \quad (2)$$

where T = torque (oz-in.)

N = number of brushes in the assembly

F = force exerted by brushes (ounces)

R = radius of the slip ring (inches)

μ = coefficient of static friction, usually about 0.1 to 0.3

Precise instrument slip rings using wire wipers require a preload force of 3 to 15 grams. Heavy-duty slip rings use preloads as high as 15 psi of wiper area.

10. Life varies from 1000 to 10,000 hours because of differences in speed, current, and noise requirements. Unlike a bearing failure or a broken brush, slip ring failure is usually not catastrophic; it appears as a gradual increase in noise level or contact resistance that eventually makes the output signal unusable.

11. Environmental—slip rings can be designed to meet almost any environmental situation, provided that cost is no problem. Hermetically sealed units ensure a measure of reliability that is usually worth the extra charge.

6.3.1. Basic Design

Slip rings have two basic designs: drum and pancake configurations. Drum slip rings are fastened to the periphery of a cylinder. Pancake rings consist of rings with various diameters mounted concentrically on a flat insulating support plate (Figure 1). The drum design has a long, thin envelope where all rings turn at a uniform rate and each ring can easily be changed to accommodate different current, voltage, and noise requirements. The ring geometry ensures minimum rubbing velocity and, consequently, long life. Good bearing support and simultaneous machining of rings results in minimum eccentricity.

Drum designs may be mounted on their own shaft and coupled to the driving mechanism or mounted directly on another rotating device (Figures 2 and 3). Most designs have the brush assemblies tangent to the periphery of the rings. A very useful variation is the so-called inside-out design where the brush assembly is inside the slip ring. This provides a very small envelope for a large number of circuits. However, the big advantage is that the entire assembly can be easily sealed against contaminants. The design and adjustment of brushes are more demanding but are well debugged by this time. Some manufacturers refer to this as a capsule design. Colletron Corporation of New York lists the following sizes for capsule rings:

length:	0.300 + 0.200 in. × (number of rings)
diameter:	up to 12 circuits 0.250
	13–30 circuits 0.375
	31–50 circuits 0.500
	51–80 circuits 0.635
	81–110 circuits 0.750

The last size is intended for signal use only.

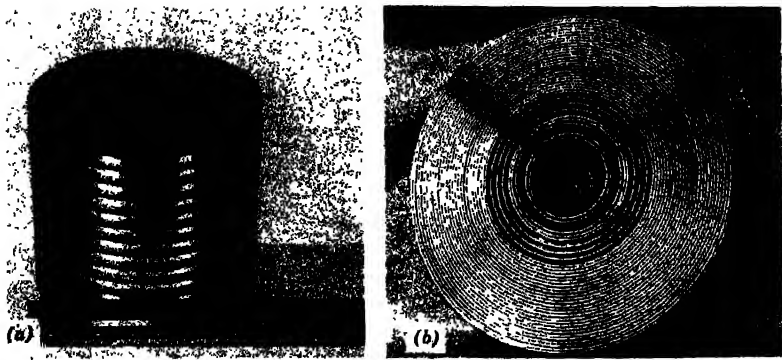


Figure 1. Drum and pancake-type slip rings. (a) Drum slip rings have the rings stacked vertically. (b) Pancake-slip rings have the rings mounted concentrically on a flat insulating support plate. (Courtesy of Electro-Miniatures Corporation. From Reference 2.)

Pancake-type slip rings are mounted in pairs above and below insulating rings which, in turn, are separated by spacers. The chief advantage of this technique is high packaging efficiency; 15 to 20 rings can be mounted per linear inch of the drum. Some designs also permit more rings to be added easily when the application changes. The pancake design results in unequal wear of the various rings due to different velocities. This fact is used to

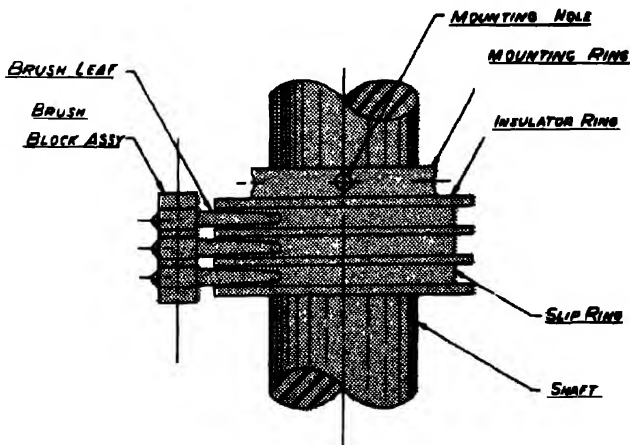


Figure 2. Drum-type slip ring assembly. (Courtesy of Englehard Industries. From Reference 4.)

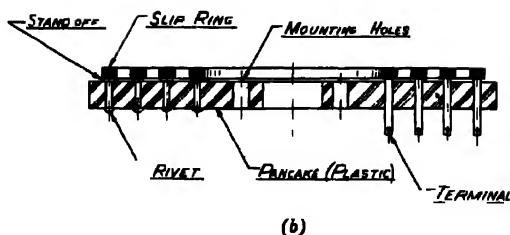
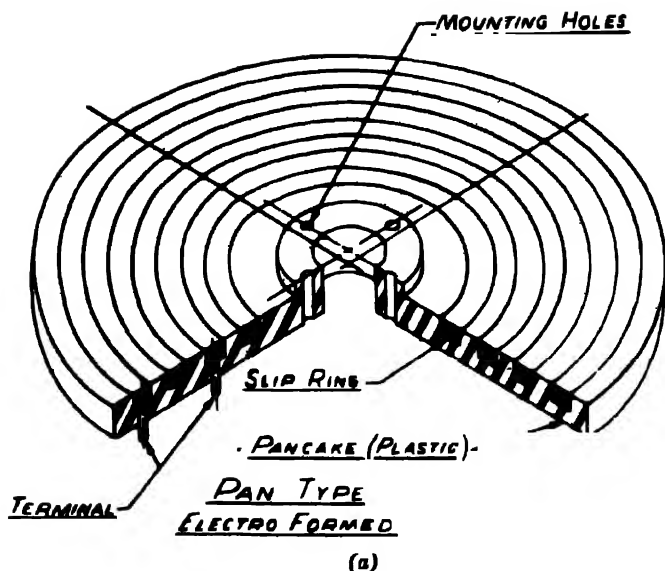


Figure 3. (a) Pancake-type slip ring plate. (b) Pancake-type slip ring assembly. (Courtesy of Englehard Industries. From Reference 4.)

advantage by placing the most critical circuits closest to the center of the disk where the speed and wear are lowest. Not only does this design provide high-packing density, but it also has the potential for adding additional plates containing 20 to 30 additional rings. Pancake slip rings are useful where the length of the supporting shaft must be held to a minimum.

6.3.2. Brush Design

Brushes provide electrical contact between the stationary and rotating members of the slip ring assembly. A typical assembly may have an aluminum housing, an aluminum or steel shaft, steel bearings, and epoxy insulation. Although the expansion coefficients are different for each material, the brush assembly must maintain contact over a wide temperature range without significant changes in preload. Two generic types are used: piston and cantilever designs (Figure 4). Piston-type brushes are loaded axially by coil springs and are generally used for high-current applications. They tend to bounce and chatter more than cantilever types but have a longer useful life. Most applications require two piston brushes per ring. The cantilever design generally is used for low to medium current applications, has low contact bounce characteristics, but more limited life. A typical design uses contact buttons riveted or welded to a support leaf that acts as the cantilever member. Shunt wires from the contact to the base of the leaf reduce overall brush resistance. Instruments normally dispense with the contact and use a single strand of wire as the contact and leaf. This results in smaller brushes, lower friction, and lower longevity.

For each type of brush, preload is critical. Some of the factors influencing brush pressure are as follows:

1. *Peripheral speed*—high rotational speed accentuates the effect of any eccentricities in the rotor and requires higher preload on the brushes. This is particularly important when the forcing frequency developed by the eccentric member approaches the natural frequency of the brush.
2. *Atmospheric conditions*—higher spring pressures are required to keep brushes clean in polluted atmospheres; a sealed unit can be used to good advantage here.
3. *Surface roughness*—to follow a rough surface, a higher than normal preload must be provided; the higher spring pressure provides the brush with a higher natural frequency. This procedure normally is encountered only with very old and abused units.

The materials used in brush assemblies are determined by the application. Slip rings for instruments and similar low current service are normally equipped with brushes that are simply strands of a noble metal. Although

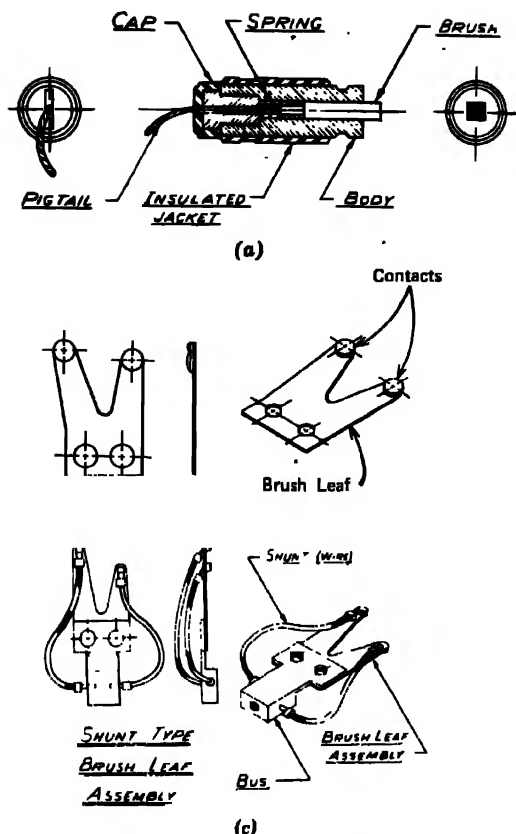


Figure 4. Piston and leaf-type brushes. (a) Piston-type brush. (b) Leaf-type brush. (c) Shunt-type leaf brush assembly. (Courtesy of Englehard Industries. From Reference 4.)

various alloys of gold, silver, and platinum have been used, most manufacturers now employ an alloy called Paliney 7, a patented alloy of the J. M. Ney Company. It is an alloy of palladium, silver, gold, platinum, copper, and zinc. It is basically a spring-type material with high resistance to corrosion, high hardness, and strength. In the annealed condition, its ductility permits easy drawing, rolling, stamping, or forming. After an age-hardening heat treatment, the alloy becomes very hard and strong. The material is available in sizes from 0.0005 to 0.020 in. diameter (Figure 5). The strand of wire serves both as a contact and as a cantilever spring. Some of the other alloys are listed in Figure 6.

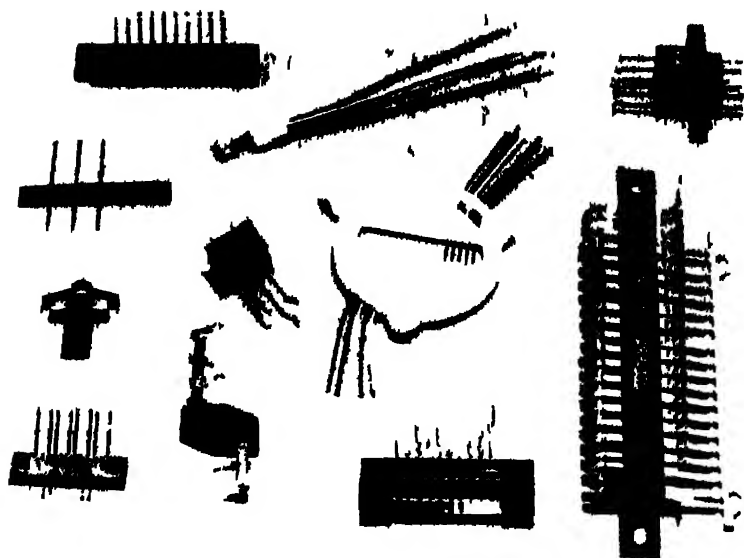


Figure 5. Brush assemblies. (Courtesy of J. F. Ney Company. From Reference 3.)

Brushes used for medium- and high-current service consist of a separate contact and some form of spring. The cantilever and plunger designs are shown in Figure 7. Three alloys dominate the field: the silver alloy group, silver cadmium oxide, and silver graphite. The silver alloy group includes such materials as silver plus copper, cadmium, nickel, zinc platinum, palladium, and iron. The high electrical conductivity of silver enables it to carry large current without excessive voltage drops. High thermal conductivity makes it possible to dissipate rapidly the heat generated when arcing occurs. Although silver oxidizes easily, these oxides decompose at relatively low temperatures incurred during arcing and revert back to pure metallic silver so that low contact resistance is maintained over long periods of time. Silver is used for low to medium current densities and contact pressures. The elements alloyed with silver are used to raise its relatively low melting point, increase its hardness, reduce its tendency to react with sulfides in the atmosphere, and increase its resistance to erosion by arcing and its resistance to welding.

Silver cadmium oxide combines the low surface contact resistance of silver with the arc-quenching characteristic of cadmium. The chief disadvantage of the addition of cadmium is that the electrical conductivity and

Engelhard Alloy No.		Rockwell Hardness					
		position	As-annealed	Cold-Worked	Electrode Conductivity	Relative Corrosion Resistance	Melting Point °F
34	AG-AU	15T 57	15T 78	40	High	1750*	5.81
109	PT-IR	B 82	B 91	7	High	3280*	11.38
141	AU-AG-PT	15T 70	15T 84	10	High	2030*	8.39
158	PT-IR	15T 90	15T 95	8	High	3310*	11.34
180	AG-PT	E 43	E 91	42	Low	1778*	5.81
285	AG-CD	15T 51	15T 83	35	Low	1620*	5.36
290	AG-PD	15T 45	15T 77	58	Low	1770*	5.55
388	AG-PD	15T 63	15T 80	27	High	1830*	5.57
485	AG-CU-NI	15T 78	15T 85	75	Low	1475*	5.27
853	AG-PD	15T 44	15T 76	79	Low	1782*	5.54
877	AG-FE	F 48	F 81	80	Low	1780*	5.40
1354	AG-C	F 44	F 72	102	Low	1780*	5.43
1388	AG-CDO	F 50	F 85	85	Low	1620*	5.28
1782	AG-CU	15T 78	15T 85	84	Low	1435*	5.27
1818	AG-CDO	F 42	F 84	75	Low	1675*	5.16
2008	AG	15T 30	15T 75	104	Low	1780*	5.54
2388	AG-CDO	F 50	F 85	85	Low	1620*	5.30
2818	AG-CDO	F 42	F 84	75	Low	1675*	5.16
3084	AU	Brinell 25	Brinell 58	74	High	1945*	10.18
3828	AU-AG	15T 50	30T 50	17	High	1815*	8.41
5512	AU-AG	Brinell 34	Brinell 87	18	High	1875*	7.54
8087	PD	15T 82	15T 78	18	High	2630*	8.41
7278	PD-AG	15T 72	15T 81	4	High	2426*	5.98
8081	PT	15T 80	15T 75	15	High	3225*	11.30
8828	AG-CD-NI	15T 50	15T 85	31	Low	1580*	5.43
10008	AG-CU	15T 70	15T 83	85	Low	1615*	5.43
10527	AG-MG-NI	15T 58-70	30T 63-67	75	Low	1840*	5.47

*Heat treated

Figure 6. Properties of commonly used contact materials. (Courtesy of Raisedlay Division, Engelhard Industries. From Reference 4.)

melting point are decreased. This alloy is used to increase the ratings of contacts normally equipped with silver alloy contacts. The cadmium oxide is normally held to about 5 to 10% of the alloy. It remains as discrete particles uniformly distributed throughout the silver matrix, and each component retains its individual characteristics.

Silver graphite combines the characteristics of silver with the good

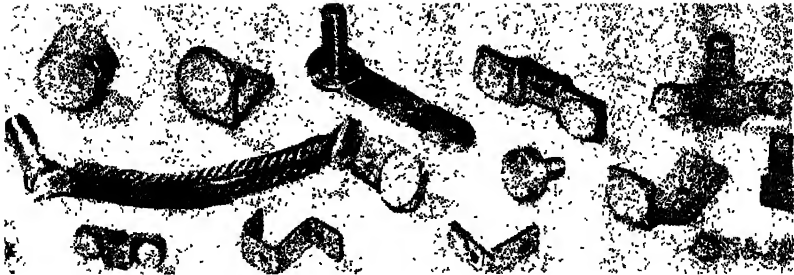


Figure 7. Silver semirefractory assemblies. (Courtesy of P. R. Mallory Company. From Reference 1.)

lubricating properties of graphite. This reduces the tendency for galling and welding. This alloy is useful for AC and DC circuits. Some companies use this alloy as their standard brush material because of its excellent service reports.

In addition to the three primary alloys, there are dozens of others. Some of them are shown in Figures 8 and 9.

The material used for supporting the contact is also crucial. It must be readily workable, strong and must exhibit good thermal and electrical conductivity. Some of the alloys used are shown in Figure 10.

Plated beryllium copper has been a favorite in recent years because of its exceptionally low hysteresis characteristics.

6.3.3. Ring Design

Fabrication of both drum and pancake slip rings requires the same general techniques. This discussion concentrates on the drum design.

The ring assembly is composed of a series of solid metal rings separated by insulators. The combination is assembled on a mandrel that forms the core of the unit. Mating surfaces of the metal and insulator rings must be smooth and parallel to permit good interfacing. The assembly must be fabricated to ensure minimum eccentricity, since any variation in radial dimensions causes a change in brush pressure, which in turn increases electrical noise. The most important system parameter in slip rings is the signal-to-noise ratio. Selection of the proper ring material is extremely important. Noise

Material	Composition % Weight	Electrical Conductivity % I.A.C.S.	Hardness Rockwell F		Density Troy oz./in. ³	Ultimate Tensile Strength p.s.i.	
			annealed	cold worked		annealed	cold worked
Mallory D-52	97.5 Ag - 2.5 CdO	88	22	70	5.39	16,000	25,000
Mallory D-53	95.0 Ag - 5.0 CdO	84	32	76	5.35	16,000	25,000
Mallory D-53X	95.0 Ag - 5.0 CdO	80	40	74	5.45	27,000	—
Mallory D-54	90.0 Ag - 10.0 CdO	75	42	84	5.16	16,000	—
Mallory D-54X	90.0 Ag - 10.0 CdO	75	45	81	5.36	27,000	—
Mallory D-154X	86.7 Ag - 13.3 CdO	68	48	84	5.31	29,000	—
Mallory D-55X	85.0 Ag - 15.0 CdO	65	50	85	5.28	30,000	—
Mallory D-505F	95.0 Ag - 5.0 Ni	95	32	84	5.49	24,000	—
Mallory D-510F	90.0 Ag - 10.0 Ni	87	35	89	5.44	25,000	—
Mallory D-50F	85.0 Ag - 15.0 Ni	80	40	93	5.28	27,000	—
Mallory D-50	85.0 Ag - 15.0 Ni	67	50	85	5.28	—	—
Mallory D-56	70.0 Ag - 30.0 Ni	55	42	87	5.02	—	—
Mallory D-51	60.0 Ag - 40.0 Ni	44	1	92	5.06	35,000	60,000

Mallory D-511	40.0 Ag - 60.0 Ni	25	42	97	4.69	—	—
Mallory D-582F	99.75 Ag - .25 C	103	45	73	5.48	25,000	37,000
Mallory D-581F	99.50 Ag - .50 C	102	44	72	5.43	24,500	36,500
Mallory D-583F	99.25 Ag - .75 C	100	39	70	5.38	24,000	35,800
Mallory D-158F	99.0 Ag - 1.0 C	99	36	69	5.33	23,500	35,000
Mallory D-558F	98.5 Ag - 1.5 C	97	33	66	5.29	22,000	33,500
Mallory D-258F	98.0 Ag - 2.0 C	77	—	—	5.04	—	—
Mallory D-58	95.0 Ag - 5.0 C	55	—	25	4.57	—	—
Mallory D-58F	95.0 Ag - 5.0 C	75	—	40	4.66	—	—
Mallory D-1058	90.0 Ag - 10.0 C	35	—	3	3.32	—	—
Mallory D-59	88.0 Ag - 2.0 C - 10.0 Ni	70	26	64	4.94	—	—
Mallory D-63X	99.34 Ag - .41 MgO - .25 NiO	70	—	97*	5.47	—	70,000
Mallory D-57F	90.0 Ag - 10.0 Fe	90	48	81	5.40	31,000	39,500
Mallory D-157	50.0 Fe - 25.0 Cu - 25.0 Ag	21	84	94	4.49	—	—
Mallory D-64F	Silver Base	101	14	82	5.50	24,500	40,500

*Listed properties are typical values

*Air Hardened

Figure 8. Properties of silver semirefractory materials. (Courtesy of P. R. Mallory Company. From Reference 1.)

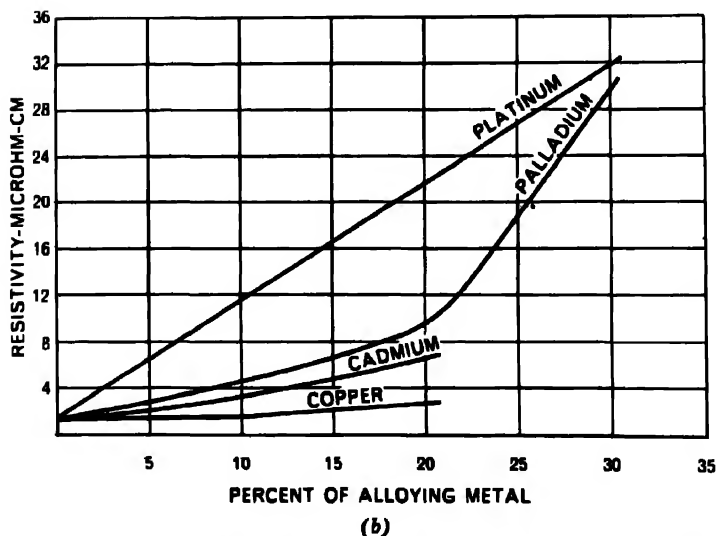
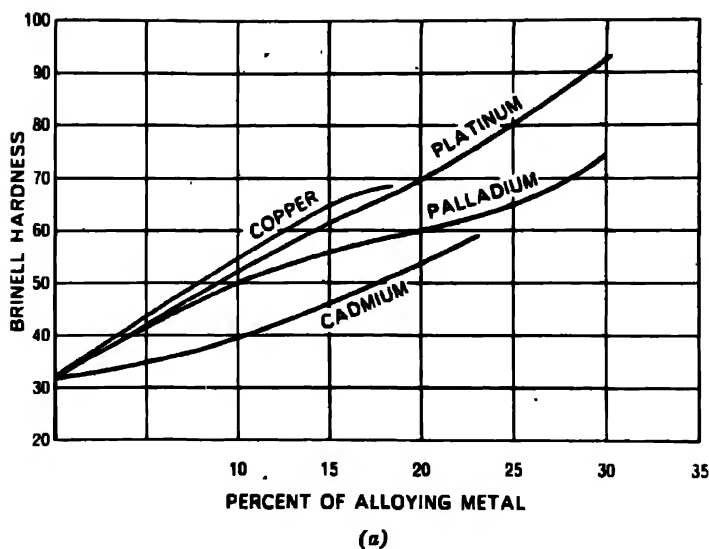


Figure 9. Effect of alloying metals on silver hardness and resistivity. (a) Increasing the percentage of alloying metals increases the hardness of silver alloys. (b) Increasing the percentage of alloying material increases electrical resistance and reduces conductivity. (Courtesy of P. R. Mallory Company. From Reference 1.)

Properties of commonly used blade, spring & backing materials

	Blade Composition % Alloy			Stress Tensile ksi (MPa)			Hardness Rockwell B			Temperature °F (°C)			% Elongation in 2"		
	Hard	Medium	Soft	Hard	Medium	Soft	Hard	Medium	Soft	Hard	Medium	Soft	Hard	Medium	Soft
OF Copper, 102	100	50	32	50	40	700-1200*	8	45	Low						
Cadmium Copper, 102	87	55	37	85	47	700-1000*	6	80	Low						
Copper, 104	80	55	45	70	38	750-1300*	4	30	Low						
Beryllium Copper, 18	45-80	75-120	50	70-80	20-45	1650-1700*	5-10	20-35	Low						
Beryllium Copper, 25	22-30	100-120 100-215	80-78	88-102 C40-45	45-78	1450-1500*	1-3	35-60	Low						
Beryllium Copper, 186	22-30	100-120 100-200	80-78	100	50	1450-1475*	1-5	35-60	Low						
Brass, 60% Zn	58	58	34	54	48	800-1450*	5	45	Low						
Common Brass, 55% Zn	44	61	37	70	53	800-1450*	5	45	Low						
Red Brass, 60% Zn	37	70	38	77	56	800-1350*	5	48	Low						
Cartridge Brass, 70% Zn	27	78	47	83	15	800-1400*	10	65	Low						
Phos Brass, 5% A, 518	15	80	48	85	28	900-1250*	8	48	Low						
Aluminum Brass, D, 614	14	85	80	86	83	1125-1850*	35	48	Low						
Cupro-Nickel, 10%, 708	8	80	44	68	15	1100-1300*	10	42-48	High						
Cupro-Nickel, 20%, 710	6.5	73	49	80	35	1100-1300*	8	45	High						
Cupro-Nickel, 30%, 715	4.8	78	55	85	40	1200-1500*	5	35	High						
Nickel-Silver, 18%, 782	6.0	85	58	87	40	1100-1400*	3	40	High						
Nickel-Silver, 16%, 778	5.5	100	80	88	50	1100-1500*	4	40	High						
Nickel-Silver, 12%, 787	8.0	85	52	86	22	1100-1800*	4	48	High						
Pewee Nickel, 388	11	150	90	C25-34	80 max	1500*	2-10	25-40	High						
Stainless Steel, 302	2.4	150 min	80	C32 min	85	1800*	12	50	High						
² Steel, 1010	17	85	48	80	80	1807-1897*	7	38	Low						
Nickel, 201	16	110	55-80	90 min	70 max	1250-1350*	2-15	40-55	High						
Monel, 400	3.5	120	75-85	83 min	73 max	1400-1500*	2-15	25-50	High						

* As Hardened
* With Copper or
* Nickel Plating
* Rockwell F

Figure 10. Properties of commonly used blade, spring, and backing materials. (Courtesy of Raisedlay Division, Englehard Industries. From Reference 4.)

is decreased by the burnishing action of the brushes on the ring surface. The degree of burnishing that can be realized is a function of ring material and the brush pressure. Twenty-four-carat gold rings burnish quickly and produce very little noise. Coin gold rings burnish almost as fast and achieve noise levels approximating those of 24-carat gold, but are much harder and serve longer. This permits higher brush pressures which further minimize the noise level. Coin gold also has a low coefficient of friction, resists oxidation, and is not susceptible to stress cracking. If the signal-to-noise ratio is less critical, coin silver and other silver alloys may be used. Various types of gold-silver alloys bridge the gap between coin gold and coin silver—18-, 14-, and 10-carat gold alloys (Figure 11). The heat developed on the ring may result in material transfer between the ring and the brush. This is called galling. Fine gold and fine silver are both susceptible to this condition. When copper is added to the alloy, the problem is greatly reduced.

PROPERTIES OF TYPICAL RING MATERIALS

ALLOY	Superficial Hardness Rockwell 15 T		Approximate ⁽¹⁾ Brinell Hardness Number		Tensile Strength 1000 psi		Specific Gravity GM/cc
	Annealed	Hard Temper	Annealed	Hard Temper	Annealed	Hard Temper	
Coin Gold ⁽¹⁾ 90% Au, 10% Cu	79-81	88-91	90-99	140-167	50-55	78-92	17.85
24 Carat Wrought Gold ⁽¹⁾	30-35	65-75	-	58-78	17-20	32-38	19.36
18 Carat Gold ⁽¹⁾ 75% Au, 15% Ag, 10% Cu	80-82	89-93	94-104	148-201	-	83-116	15.59
10 Carat Gold ⁽¹⁾ 41.66% Au, 12.4% Ag 36.17% Cu, 9.33% Zn	80-84	90-93	94-114	157-201	67-75	95-120	11.46
Coin Silver ⁽¹⁾ 90% Ag, 10% Cu	70-72	84-88	66-71	114-140	34-42	60-65	10.35
Baker #416 ⁽²⁾	-	-	-	-	40	60	16.2
Ney #90 ⁽³⁾				120-130 ⁽⁴⁾			

(1) Reprinted Courtesy Improved Seamless Wire Company

(2) Reprinted Courtesy Baker and Company

(3) Reprinted Courtesy J. M. Ney Company

(4) These hardness values are converted from other scales and the refore may be considered approximate only, since such conversions are of an arbitrary nature.

Figure 11. Properties of typical ring materials. (Courtesy of Poly-Scientific Corporation. From Reference 5.)

The shape of the rings may be flat, U-grooved, or V-grooved (Figure 12). The flat design is the least expensive, but permits considerable axial movement of brushes, thus leading to increased noise. The U-grooves capture the brushes.

Insulators must provide good insulation between adjacent slip rings as well as mechanical strength for structural integrity. Figure 13 lists the commonly used materials and some of the critical parameters. Two important properties not listed are dimensional stability and outgassing characteristics. Dimensional stability is important because any change in insulator dimensions causes the rings to shift position, which may destroy the brush-to-ring alignment. Outgassing is the evolution of gas from the material as a result of exposure to high temperatures or low pressures. For some materials neither condition is required to catalyze the reaction. The gases released tend to settle on the nearest available surface. This deposit may form on the

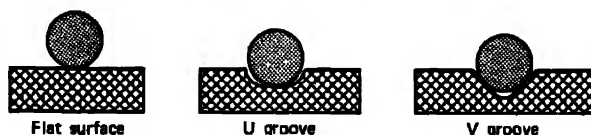


Figure 12. Typical configurations of wire brush and slip ring arrangements in common use. (Courtesy of Benwill Publishing Company. From Reference 17.)

	Tensile Strength	Impact Strength	Expansion Coefficient	Service Temperature	Dielectric Strength	Volume Resistivity	Dielectric Constant (Low f)
Unfilled epoxy	2	3	3	3	2	2	3
Unfilled rigid polyester	2	4	4	3	2	2	3
Nylon	2	2	4	3	2	2	3
TFE-Teflon	4	3	4	1	2	1	1
High pressure laminates	1	1	1	2	3	2	4
Filled phenolics	2	2	2	3	3	2	4
Diaryl pyrolytes	2	2	2	2	3	3	4
Silicones	3	4	3	1	3	2	2
Filled melamines	2	3	2	2	3	2	4

Rating: 1 = excellent, 2 = good, 3 = fair, 4 = poor

Figure 13. Properties of slip ring insulating materials. (Courtesy of Benwill Publishing Company. From Reference 17.)

slip ring and cause high contact resistance. When it accumulates on the bearings, higher driving torque results. In space applications gas deposition on radiation surfaces can lower the emissivity and cause the thermal balance to change. Optical components would suffer severely if deposits formed on the lenses.

There are three primary methods for manufacturing slip ring assemblies:

1. *Stacking*—preformed rings of conductors and insulators are assembled on a mandrel. This method is very suitable for small quantities, because no expensive tooling is required. The disadvantage is that the materials are limited to those that can be easily machined, since the final process is to machine all rings at assembly to eliminate eccentricities (Figure 14a).

2. *Casting*—the rings and the attached leads are positioned into an accurately grooved mold, which is then filled with the desired plastic. This process is ideal for large production runs (Figure 14b). Any low or medium viscosity plastic can be used. Normally there is very little machining of plastic after casting. The principal disadvantage is the cost and time required to fabricate the mold and associated casting equipment. The technique is unsuitable for high viscosity or highly filled plastic systems. Some manufacturers succeed with viscous materials by using special processes, but the cost then increases significantly.

3. *Electrodeposition*—a multistage production process that includes casting plastic around a structural shaft member and lead wires. Then the plastic is grooved for rings, a metal surface is deposited at the bottom of the groove for contact with the proper lead wire, the ring material is plated into the groove to build up the ring, and, finally, the ring is machined and surface-finished (Figure 14c). This yields the most accurate ring spacing of all methods. The main disadvantages are high tooling costs and the limitation to materials that can be reliably electrodeposited.

6.3.4. Lubrication

Wear on sliding assemblies has been attributed to four primary causes:

1. *Adhesive or cohesive action*—part of the contact or brush assembly adheres to the ring and is broken away from its parent body. The result is an increase in surface roughness for both parts. This effect is largely eliminated when a film separates the two rubbing surfaces. In some cases, the material itself forms an oxide, but in others a foreign film in the form of a lubricant must be used.

2. *Abrasion*—when two rubbing surfaces of different hardness are mated together, the softer material will wear more rapidly. Slip ring assemblies are fabricated so that the rubbing materials have about the same hardness.

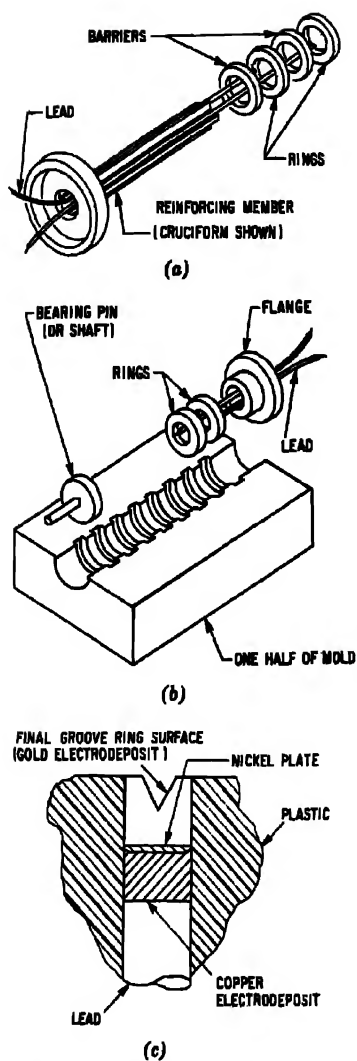


Figure 14. Methods of fabricating slip rings. (a) Stacked slip ring. (b) Plastic is cast or molded into spaces between components. (c) Electrodeposited type ring. (Courtesy of Poly-Scientific Corporation. From Reference 5.)

Since a tolerance on hardness is necessary, it is selected to allow the rings to be somewhat softer than the brushes, since they are sufficiently massive to accommodate wear better than the brushes.

3. *Corrosive action*—this problem is easily cured by proper plating.

4. *Surface fatigue*—failure as a result of repeated loading. Lubricants such as mineral oil and Kel-F-10 have been used successfully on slip rings. Space applications require a lubricant that will not outgas at low pressures; consequently, molybdenum disulphide has been added to conventional lubricants. Graphite-silver brushes do not require supplementary lubrication, but do not lubricate well above an altitude of 100,000 ft. A new arrival is a coating called barium fluoride eutectic. This material is applied by vacuum impregnation and produces a low coefficient of friction that works well at high temperatures and high sliding velocities. Typical friction coefficients reported are 0.06 at 1500°F at a velocity of 2000 ft/min. Another method of ensuring lubrication at high altitudes is to seal the assembly using labyrinth seals; this design allows a controlled leak of gas into space, but maintains sufficient pressure to prevent the lubricant from vaporizing. Sometimes a wick is provided in the system that evaporates a liquid at a sufficient rate to keep the slip ring at a safe pressure. A gas flask and bleed valve have also been used to accomplish this purpose.

Supplementary lubrication is considered mandatory for surface speeds above 250 ft/min. Prime considerations in selecting a lubricant are low viscosity and resistance to polymerization. The biggest risk in using a lubricant is that it may collect any available dust or contaminants in the atmosphere and work them into the rings and bearings with disastrous results.

6.3.5. Bearings

One item not usually covered in manufacturers' literature on slip rings is the bearings used. One reason is that many slip rings were designed for use in the customer's own journals and shafts. As the industry became more sophisticated, an increasing number of instruments were designed as self-contained units that could more easily be coupled to the user's shaft than be incorporated into it. In small and miniature slip ring assemblies, a good instrument will include an ABEC-5 or ABEC-7 class bearing. Axial clearances in the bearing affect the alignment of the brush; variations can cause excessive wear and noise. Radial bearing clearances cause changes in brush pressure, hence also affect noise and wear characteristics. Ideally, a properly preloaded bearing would eliminate all clearances and exhibit a small preload that prevents any axial shift until the external force exceeds the preload. Practically, the friction levels would be prohibitive. The usual

compromise is an axial end play of about 0.0003 to 0.0007 in. Brush preload is normally sufficient to accommodate moderate radial yield in bearings.

6.3.6. Slip Ring Noise

Electrical noise introduced by sliding contacts is probably the least understood of all slip ring operating parameters. This is partly because the noise signal is generally random in both frequency and amplitude and no exact analytical expression for the wave exists. Random waves, having no period, are probabilistic and are usually defined in terms of functions of power spectral density, autocorrelation, and amplitude probability density. To simplify slip ring specifications, industry has ignored the principles of random wave analysis and has resorted to more convenient direct voltage measurements.

The noise developed by sliding contacts is caused by fluctuations in contact resistance as motion occurs. Resistance changes derive from the variations in contact area and pressure as a result of ring eccentricities and contamination. Slip ring noise is measured with an oscilloscope connected across the slip ring circuit, through which a known direct current is flowing. The maximum peak-to-peak voltage fluctuation over an operating cycle of the slip ring is defined as the noise. The measured value is usually converted to resistance units for convenience. The measurement is expressed in milliohms. Normal instrument slip rings will have noise of about 1 to 10 m Ω . It tends to increase under conditions of shock, vibration, and dynamic stress. Noise also becomes greater as surface speed increases and as the contact surfaces wear or become contaminated. Multiple brushes reduce noise appreciably. If we assume that noise is a resistive phenomenon, then measurement of noise is reasonably free of current limitations. The current used in tests should be large enough to permit the utilization of conventional test equipment and to be representative of the final application.

An intermittent circuit is defined as one in which continuity is temporarily interrupted. A random spike, as observed on an oscilloscope, is one that exceeds the average noise level by a factor of 5 or more. It is generally caused by contact bounce or severe contamination. If the period of a spike is critical, it may be detected by an SCR circuit and measured on an oscilloscope.

6.3.7. Instrument Slip Rings

Instrument slip rings are typically cantilever types. The brushes are made of round or flat wire and are encapsulated in a relatively stiff block. The

convex shape at the end of each brush is designed to maintain line contact with the slip ring. The cross-sectional area of the brush wire is designed for the use of about 4 mA per circular mil for noble-metal wires. Much higher ratings are sometimes employed at the expense of life. The principal failure mode of brushes is overheating; consequently, any ventilation or heat-sinking techniques will extend the life of the slip ring. Some of the designs use more than one brush per ring to reduce the current per brush and also to ensure that under vibration or shock environments at least one brush is always in contact with the ring. A circular brush in a V-shaped ring is one method of easily obtaining two points of contact. The limiting factor on brush spacing is the dielectric strength of air. The design value used for calculating separation is about 50 V/mil for reasonably clean air; for vacuum or high-altitude work the figure is only about 0.5 V/mil. The brush block materials have superior insulation resistances of 200 to 400 V/mil. They are also the principal medium for transferring heat from the brushes to the ultimate heat sink. Although their heat conductivity properties are poor, they are nevertheless much superior to air, the only other medium available for heat transfer. Radiation does not play a significant role in cooling the brush block. Flexure of brushes under shock and vibration must also be considered. The preload per brush is generally a compromise between excessive frictional torque and contact prevention bounce. On small brushes the preload range is about 5 to 15 grams. Brush surface finishes of 4 to 8 $\mu\text{in.}$ are usually required.

6.3.8. Power Slip Rings

Power slip rings use plunger brush designs as well as highly preloaded cantilever designs. The plunger design is preferred for current ratings above 100 A; the cantilever design is more economical in the 20 to 100 A range. The plunger-type brush is a circular or square rod of silver graphite that is spring-loaded. The allowable current density is about 300 A/in.². The preload on power brushes is about 10 to 15 psi. Silver-alloy rings are used. The chief method of cooling the brush is through the body of the brush and with rotational windage. Leaf-type brushes are characterized by good-bounce performance and shorter life than plunger types. The contact is a button-shaped alloy fastened to the end of a relatively thin cantilever leaf made of beryllium copper. Proper heat treating and forming of the leaf provides a reliable preload under a wide range of temperatures. This alloy is selected because of its good electrical conductivity, low hysteresis, and good thermal conductivity. The leaf must be designed as a cantilever beam with a given natural frequency. Consequently the cross-sectional area is

restricted; hence the current capacity is also limited. Normal practice on high-current brushes is to provide short wires from the base of the leaf to the contacts so as to form a path around the leaf to prevent overheating. Determination of the preload is again a compromise between low contact resistance and low noise (high preload) and long service life (low preload). For a given current the optimum preload is usually determined empirically. On power applications, four or more brushes per ring is accepted practice.

6.3.9. Radio- and Video-Frequency Slip Rings

Slip ring assemblies for communication equipment have been used up to frequencies of 100 mHz at very low signal levels. The instruments must have a very wide frequency range with flat characteristics. The radio/frequency (RF) slip ring is part of the terminating circuit of a transmission line. If the slip ring assembly is inserted in a matched circuit, any standing wave is due to the presence of the slip rings. If the impedance the line "sees" is the characteristic impedance, then the slip ring impedance is the characteristic impedance of the line and there is no standing wave present. Energy transfer is maximum. When analyzing slip ring parameters, they may be considered a lumped circuit up to frequencies of 60 mHz. Slip rings should be considered distributed circuits if the dimensions of the ring exceed 0.05 of the wavelength. The terminating impedance of the transmission line is essentially a T-network having inductive and capacitive elements with the characteristic impedance as the load. By calculating the terminating impedance, if the characteristic impedance of the line and the reflection coefficient are known, the voltage standing wave ratio can be obtained. Because of the use of low loss materials, insertion loss usually is 0.1 dB or less, excluding cables.

The major factor in limiting cross talk between rings is the space between the fixed shield and the rotating shield (Figure 15). This distance should be kept to a minimum and the axial length as large as is practical. The usual gap length-to-width ratio is about 6-1. The shield ring is grounded to the housing by brushes at several places to ensure that it is kept at ground potential. By proper design, it is possible to limit cross talk between rings to 80 dB at 150 mHz and to 50 dB at 60 mHz.

Power-pulse slip rings are used to improve the weight distribution of a radar system. The design of these elements involves all the features of a high-voltage slip ring as well as the impedance characteristics of an RF component. Special attention must be given to the dielectric materials used to prevent arc-over. Since ozone is generated more readily in high-voltage discharges, its effect on materials must be carefully evaluated. The slip ring

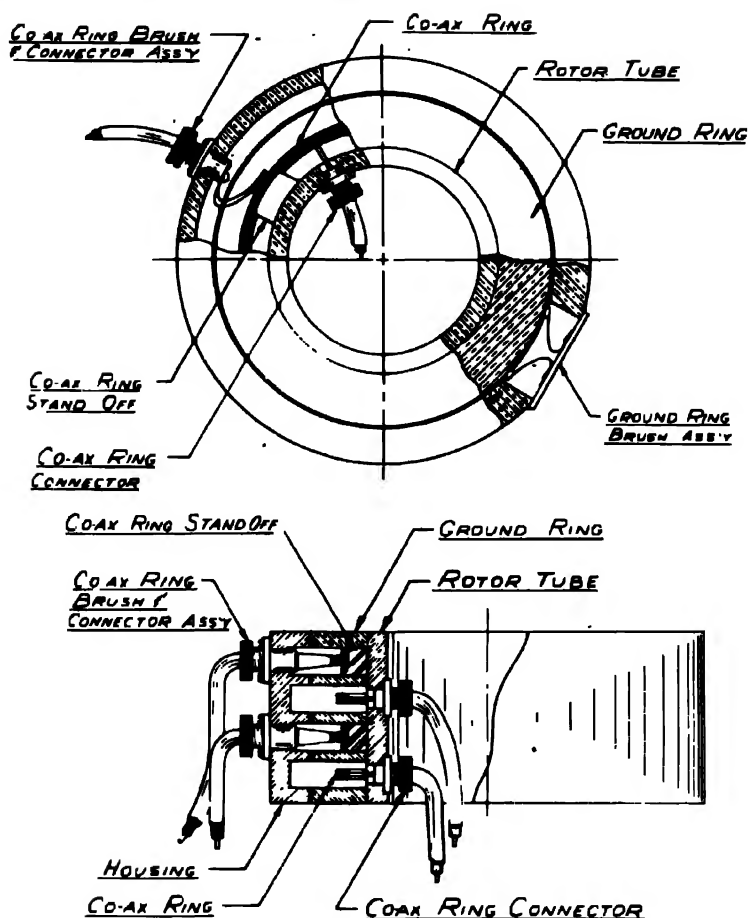


Figure 15. Radifrequency slip rings. (Courtesy of Englehard Industries. From Reference 4.)

must transmit the current pulse without distortion; since the rise time of the current pulse is affected by circuit capacitance, the time constant of the slip ring must be much less than the rise time of the current pulse.

6.3.10. Environmental Considerations

Temperature variations produce the most severe effects on slip ring assemblies. Thermal variations alter bearing clearances and preloads, which

in turn affects noise and torque levels. The following parameters should be considered carefully for designs intended for critical applications:

1. The coefficient of expansion of dissimilar materials may cause misalignment, which can seriously increase noise. The mounting surfaces are also potential trouble areas.
2. The distortion point of plastic insulators should be considered against the background of heat generated without the cooling effects of air (space applications).
3. The outgassing of nonmetallic elements at high altitudes is an ever-present danger.
4. The creep of solder joints at temperatures approaching 300°F may be a source of circuit failures.
5. The increase in circuit resistance at elevated temperatures will increase parasitic heating effects.

As atmospheric pressure is reduced, cooling by convection becomes impossible and all resistive elements must be provided with heat sinks. Current ratings are normally reduced by 20 to 50% under these conditions.

Lubrication at low pressures must be performed by compounds with very low vapor pressures. Dry lubricants are preferred. Oxides normally forming on contacts that assist lubrication do not form at high altitudes.

Relative humidity above 70% increases the moisture absorption of plastics, thus lowering their volume resistivity and dielectric strength. The excess moisture may combine with minute traces of processing salts to produce surface leakage paths. High humidity followed by subsequent rapid cooling may result in condensation of moisture on electrical contacts. This is another good reason for using hermetically sealed units.

Most good slip ring assemblies are designed to withstand the effects of shock and vibration. The forcing function that corresponds to the natural frequency of the brush assemblies is the most dangerous. The biggest problem is to determine which type of forcing function is present in a particular type of equipment and environment. The design of the unit is relatively simple once this information is available.

6.4. POTENTIOMETERS

The classical application of potentiometers has been as voltage-measuring instruments. In contrast with rheostats, they are not designed to consume power. The basic potentiometer measures an unknown voltage by opposing it with an equal and opposite voltage (Figure 16). The characteristic that makes this device useful is voltage division. All applications of potentiometers (commonly called pots) are derived from this property. Usage is virtually

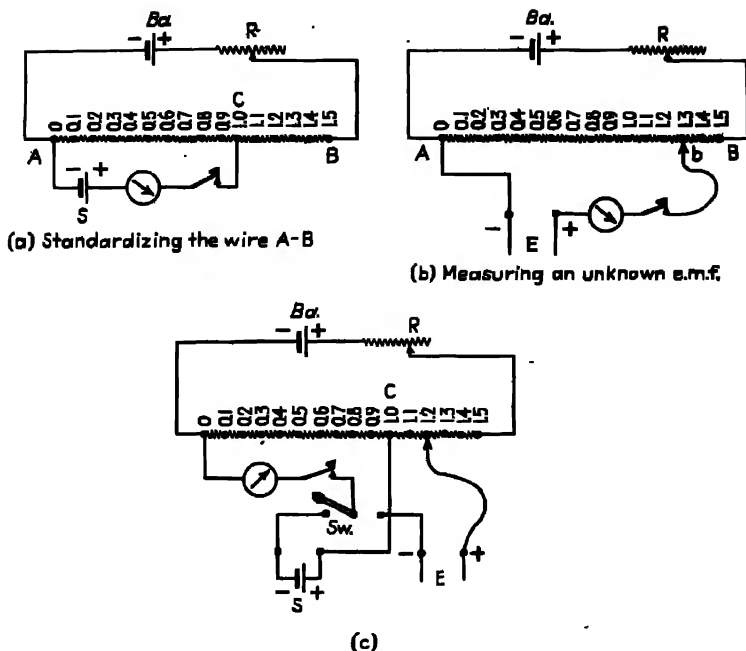


Figure 16. Simple potentiometer. The two circuits (a) and (b) may be combined into one by the use of a SPDT switch. When the switch is in its left-hand position the standard cell is in circuit for calibration (a). When the switch is in its right-hand position, the unknown EMF is placed across the wire *AB* so that its value may be determined. (Courtesy of McGraw-Hill Book Company. From Reference 18.)

unlimited and includes transducers, servosystems, analog computers, bridges, and most primary measurement instruments.

6.4.1. Classification

The array of pots available to the engineer is tremendous. They vary from multipurpose instruments to units useful in only one segment of a specialized instrument. A generally accepted classification is by function:

1. Single-turn pots.
2. Multiturn pots.
3. Ganged pots.
4. Trimmer pots.
5. Nonlinear pots.

Single-turn pots are the units most in use. They are the most likely instruments to be available for a specific resistance, power rating, and price. Selection of single-turn pots extends from "beer to champagne" budgets.

Multiturn pots are used to obtain the maximum resolution built into the instrument. The basic resolution is the spacing between adjacent turns in the resistance winding. For example, if a 1000- Ω pot contains 1000 turns of wire, the basic resolution is 1 Ω . If this winding is built into a single-turn pot, the user must be capable of turning the control knob 1/1000 of a turn to obtain increments of 1 Ω . When the same winding is put in a 10-turn instrument, the user must turn the knob 1/100 of a turn to change the setting by 1 Ω . This is the chief advantage of multiturn instruments. There are no gains in linearity or accuracy. These devices are more costly than single-turn models, since some form of gearing must be provided. Some multiturn units are longer than the corresponding single-turn units. This is to provide space for a longer winding and the gearing assembly. The longer envelope makes heavier gage resistance wire feasible without loss of resolution. Since the mass supporting the coil is relatively greater, heat dissipation is better and the power rating is increased. The gear train also results in lower input torques. The chief disadvantages of this product are cost, increased complexity, and unsuitability for low resistance values.

Ganged potentiometers are individual units mechanically coupled so that all wiper arms turn simultaneously and the output position of each unit is identical. This method is used when one sensor controls several motors or circuits that must be in phase with one another. Pots designed for ganged assemblies can be stacked almost without limit. The ultimate consideration is the torque available to drive them.

Trimmer pots are used as verniers on the settings of other resistive elements. Their resistance range is generally small but extremely precise. The most important characteristics are high resolution and stability. Sometimes they function as variable resistors rather than voltage dividers. The instrument generally is available in a small rectangular envelope with some form of lead screw drive. Ten- and twenty-turn units are common. The design problem associated with trimmer pots is that they cover a small range of resistance, which implies that the resistive element wire must be relatively large. However, large-diameter wiring results in poor resolution. The compromise is usually a unit a little larger than ideal but compatible with system requirements. Trimmer potentiometers are normally adjusted with a screw driver; consequently, input torque is not a problem. This enables the manufacturer to concentrate on designs that are immune to the effects of vibration, shock, and thermal variations.

Nonlinear potentiometers are used when the output of the instrument must be proportional to a sine, cosine, tangent, or other nonlinear function.

The accuracies obtainable with these devices are not so good as those from linear units. Resolution is different on various sections of the curve (Figure 17). The ratio of maximum slope to minimum slope must be limited to 4-1 or 5-1 if the function is to be practically obtainable. Some of the following techniques are used to produce nonlinear functions:

1. *Varying wire spacing.* Consider a number of turns of wire wound on a cylindrical mandrel with the pot wiper arm moving along the axis of the mandrel. If the turns of wire are close together, a given displacement of the slider results in a large change of resistance; if the coils are spread farther apart, the same movement of the slider causes less change in resistance. By carefully varying the pitch of the coils to correspond to a given function, nonlinear outputs can be achieved. This technique requires special coil winding equipment.

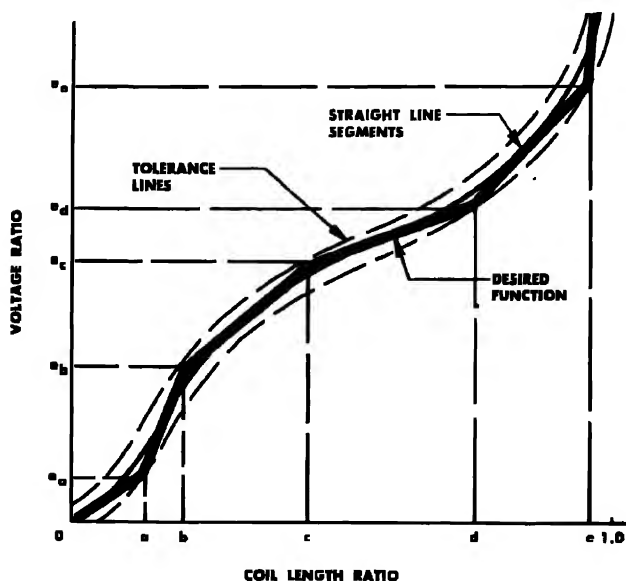


Figure 17. Typical nonlinear function. (Courtesy of Conrac Corporation. From Reference 9.)

2. *Varying the diameter of the mandrel or the shape of the card.* If resistance wire is wound on a conical mandrel, the output of the pot varies as a function of the slope of the cone. If the shape of the mandrel is changed irregularly,

the pot output follows the same form. This technique is easier when the wire is wound around a card rather than a mandrel, since the change can be made by simply cutting steps or slots in the card.

3. *Resistance shunting.* This method involves placing a number of taps on a resistance coil and adjusting the slope of each segment of the coil by shunting it with an appropriate resistor (Figure 17). This results in a series of straight-line slopes that can be used to approximate almost any curve. Some functions require combining resistance shunting with varying the mandrel shape.

4. *Ganging potentiometers.* By wiring two or more pots in series and starting the point of actuation at various points, we achieve the effect of superimposing a series of straight lines. The basic pattern is an increasing nonlinear function. This method is specially adaptable to power-law functions.

5. *Varying coating thickness.* In film-type pots the thickness of the coating can be varied to produce nonlinear effects. The most popular nonlinear pots produce sine-cosine functions. This type consists of two wipers spaced 90° apart which simultaneously generate sine and cosine functions. The output of the pot is switched from one to the other every 90° ; the negative portion of the curve is obtained by switching polarity. Figure 18 illustrates the various types of nonlinear function pots.

Another way of categorizing pots is by the type of resistance element:

1. Wirewound.
2. Cermet.
3. Hot-mold carbon.
4. Carbon film.
5. Thin-metal film.

A short description of each type of unit was given in Section 2.2.1. We now discuss them in more detail (Figure 19).

Wirewound potentiometers have been the standard of precision for many years. Generations of engineers have used and abused them and have tended to retain them long after more efficient materials became available. The outstanding virtue of wirewound units today is their unequalled temperature coefficients of resistivity. By selecting such wire materials as Karma, Advance, or Manganin, thermal resistive effects can be held to practically zero. Environmental characteristics of wirewound units are also good, but not superior to those of other materials. On the negative side, electrical noise is random and nonrepeatable and usually increases during the life of the pot because of wear and oxidation. Nonwirewound units tend to remain repeatable and unchanged during usage. From a systems standpoint, the wirewound pot must be considered to be an *RLC* component. The inductance of the coils and the capacitance associated with adjacent turns

304 Slip Rings and Potentiometers

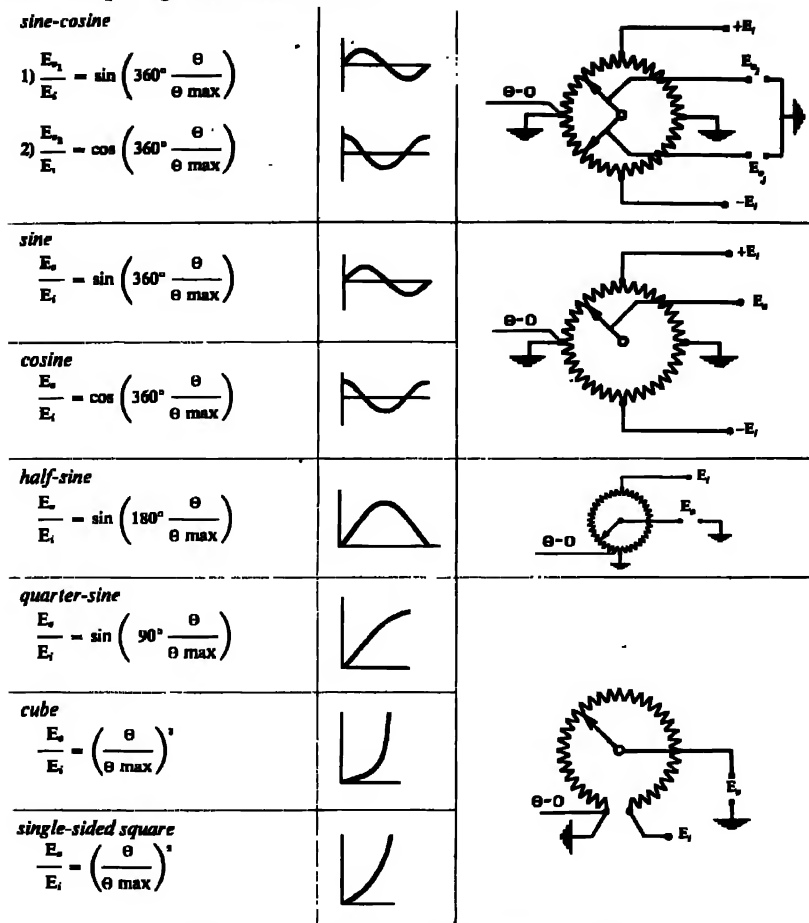





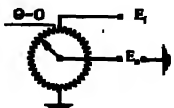


Figure 18. Nonlinear potentiometer functions. (Courtesy of Beckman Instrument, Inc. From Reference 16.)

become significant as the frequency increases. Resolution of wirewound units is limited by the thickness of the wire.

Cermet is a ceramic-metal substance that is a relative newcomer to the nonwirewound pot field. As in all homogenous resistive materials, its resolution is theoretically infinite. Its environmental stability and life are second to none. Temperature-resistive effects are somewhat higher than in wirewound units. In contrast to wirewound units, it is ideal for AC usage with applications up to 200 MHz. Its resistance range is greater than that of any other material—from 10 to 2 MΩ in standard units. Cermet is excep-

<i>tangential</i> $\frac{E_o}{E_i} = K \tan \left[\phi \frac{\left(\theta - \frac{\theta_{\max}}{2} \right)}{\left(\frac{\theta_{\max}}{2} \right)} \right]$		
<i>log (exponential)</i> $\frac{E_o}{E_i} = A^{\left(\frac{\theta}{\theta_{\max}} - 1 \right)}$		
<i>double-sided square</i> $\frac{E_o}{E_i} = \left[\frac{\left(\theta - \frac{\theta_{\max}}{2} \right)}{\left(\frac{\theta_{\max}}{2} \right)} \right]^2$		

output voltage = E_o

input voltage = E_i

shaft rotation degrees = θ

maximum electrical

function angle = θ_{\max}

maximum tangent

angle of series = ϕ

constant dependent on $\phi = K$

number of log cycles = b

Figure 18—continued

tionally good at surviving overload surges. The price of cermet pots is moderate.

Hot-molded carbon is a mixture of carbon and thermosetting plastic binder that is fabricated by molding. The chief advantages of this material are infinite resolution and low cost. Unfortunately, this design does not resist well the effects of humidity and environmental stress. The temperature coefficient is also high and negative. For applications where long rotational life under dry conditions is required, hot-molded carbon is a good choice. This material is still being improved and may become free of environmental limitations in the near future.

CHARACTERISTICS	WIREWOUND	NONWIREWOUND TRIMMERS			
		Cermet	Hot-Mold Carbon	Carbon Film	Thin-Metal Film
Setting Ability	Poor to Good	Excellent	Excellent	Excellent	Excellent
Resistance Range	Low to Medium	Low thru High	Medium to High	Medium to High	Low to Medium
Power Rating	Medium	High	Low	Low	Medium
Temp. Coefficient	Lowest	Low to Medium	High	High	Low
Environmental Stability					
High Temperature	Good to Exc.	Excellent	Poor	Fair	Excellent
Load Life	Good to Exc.	Excellent	Fair	Fair	Excellent
Humidity	Good to Exc.	Excellent	Poor	Poor	Excellent
Rotational Life	Good	Excellent	Good	Good	Fair
Rheostat Usage	Good	Fair	Poor	Poor	Fair
AC Usage	Fair	Excellent	Excellent	Excellent	Excellent

Figure 19. Potentiometer characteristics. (Courtesy of Beckman Instrument, Inc. From Reference 16.)

Carbon film pots are composed of a thin deposited coating of carbon on a nonconductive base. The thickness of the carbon determines the resistance of the unit. As in all nonwirewound units, the resolution is excellent. Environmental performance is better than hot-molded carbon, but not so good as in wirewound or cermet pots. An interesting characteristic is that the failure mode is gradual; the wearing of the film produces increasing readings at a particular weak point rather than a catastrophic failure. Rotational life is lower than in hot-molded carbon, cermet, or wirewound units. Price is average. Power rating is low. Linearity is excellent and comparable to wirewound units—values as low as 0.01% are available.

Metal film pots are composed of a thin, vapor-deposited layer of metal on ceramic or glass. This material has excellent resistance to environmental conditions and low thermal coefficients. The metal provides a path along which the heat generated at the wipers can be conducted uniformly to the entire resistive surface. The heat is also transmitted to the ceramic base. Although ceramic is not a good conductor, it can be considered as a heat sink for the metal film, since it is relatively massive and the heat is uniformly distributed over its entire boundary with the metal. This technique permits use of metal film pots in the 200 to 250°C range, which is certain to be extended in the near future. Rotational life is rated to be only fair, because the metal wiper arm rubs against the metallic resistance film. If the film is sufficiently thick (low-resistance pots), the problem is identical to that of a slip ring and the same techniques are applicable. Linearities of 0.5% or larger have been reported by various manufacturers.

A final consideration when selecting a pot is the possibility of catastrophic failure, that is, an abrupt open circuit. Wirewound units are the most susceptible to this difficulty, and metal film pots are almost as susceptible. The other nonwirewound units rarely fail in this way. This comparison is hardly comprehensive, however, since good wirewound pots rarely fail at their resistive elements. Broken terminals due to carelessness or abuse are much more common, and this can happen to any type of pot.

6.4.2. Construction

Construction of potentiometers is a compromise between system requirements and commercial realities. Successful designs are based on units that can be mass-produced rather than being manufactured under conditions where time and money are no object. For these reasons, pots are one of the most thoroughly researched components in existence. Every company has developed special methods that provide some technical or commercial advantage for its line of products. This section discusses only some of the generic features of potentiometer construction and their system implications.

Wirewound units are emphasized, since they best illustrate the physical principles of potentiometer design.

The main parts of a potentiometer are the following:

1. Housing or base.
2. Resistance element.
3. Slider contact.
4. Support shaft and bearings.
5. Terminals.
6. Enclosure.

The housing supports the resistance element and provides guidance for the support shaft (Figure 20). It holds the resistance shaft in a fixed

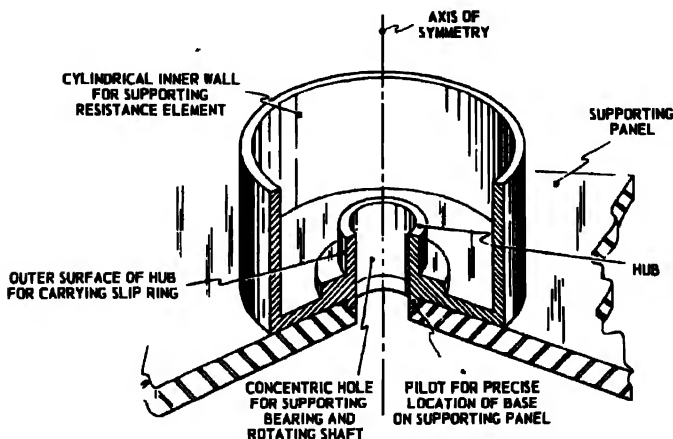


Figure 20. Potentiometer base cut away to show its precision surfaces. (Courtesy of Technology Instrument Corporation. From Reference 10.)

relationship with respect to the axis of the pot. It also supplies a heat sink for the winding. Aluminum housings possess excellent thermal characteristics as well as precise mounting diameters. Molded plastic housings are used where cost is the prime consideration; they also excel at high-frequency operation when shunt capacitance between the housing and resistance winding must be minimized. Dimensional and environmental stability in plastics is generally inferior to that in metal units. However, modern plastics technology is producing some exceptions to this rule in certain categories.

The resistance element is the heart of any pot. Some of the wire used is shown in Figure 21. The important considerations in resistance wire are

Alloy Name	Alloy Composition, Approx., per cent	Resistivity, ohms per circular mil foot at 77 F (25 C)	Mean Temperature Coefficient of Resistivity, ppm per deg. Cent. based on reference temp. of 77 F (25 C)	Maximum Thermal emf versus Copper, mv per deg. Cent.	Temperature Range (for Values in Columns 6 and 8), deg. Cent.	Specific Gravity	Coefficient of Thermal Expansion $\times 10^{-6}$ per deg. C 30 C to 300 C
KARMA*	76 Ni, 20 Cr + Fe + Al	800	0, \pm 5	+0.003	-65 to +150	8.10	13.3
NICHROME* V	79 Ni, 20 Cr, 1 Si	650	+ 80, \pm 20	+0.006	-65 to +250	8.41	12.5
NICHROME	79 Ni, 20 Cr, 1 Si (Stahlrad, 808 Alloy)	675	+ 60, \pm 20	+0.006	-65 to +150	8.41	12.5
NICHROME	58 Ni, 16 Cr, 1 Si, bal. Fe	675	+ 140, \pm 30	+0.002	-65 to +250	8.25	12.8
ADVANCE*	55 Cu, 45 Ni	394	0, \pm 20	-0.045	-65 to +150	8.90	14.5
Manganin	84 Cu, 12 Mn, 4 Ni	290 ⁽¹⁾	0, \pm 15	-0.003	+15 to +85	8.41	10.7
MIDORM*	77 Cu, 23 Ni	180	+ 180, \pm 30	-0.037	-65 to +150	8.90	15.7
HYTEMCO* ⁽¹⁾	70 Ni, 30 Fe	120	+3900, \pm 300 +4500, \pm 400	-0.04 -0.04	-50 to +30 +20 to +100	8.46	12.0
95 Alloy	90 Cu, 10 Ni	90	+ 450, \pm 50	-0.026	-65 to +150	8.90	16.0
LOHM*	94 Cu, 6 Ni	60	+ 700, \pm 200	-0.022	-65 to +150	8.90	16.2
30 Alloy	98 Cu, 2 Ni	30	+1400, \pm 300	-0.014	-65 to +150	8.91	16.4

	Specific Heat Grain Calorim	Thermal Conductivity Watts per CM per deg. C. at 100 C	Approx. Melting Point deg. C.	Tensile Strength at 20 C Pounds per Sq. In.		Pounds per Cubic Inch
				Max.	Min.	
Karma.	.104	.130	1400	180,000	130,000	.293
Nichrome	.107	.132	1350	200,000	95,000	.298
Nichrome V	.106	.112	1400	200,000	100,000	.294
Advance	.094	.212	1210	100,000	60,000	.322
Manganin (Wire only)	.097	.190	1020	90,000	40,000	.304
Midorm	.092	.150	1100	100,000	50,000	.322
Hytemco	.125	.289	1425	150,000	70,000	.306
95 Alloy	.092	.605	1100	75,000	35,000	.322
Lohm	.092	.907	1100	100,000	50,000	.322
30 Alloy	.092	1.62	1100	60,000	30,000	.321

Figure 21. Properties of alloys used as resistance wire. (1) the values given either meet or surpass the specified requirements of ASTM Specification B267-65T. (2) Temperature coefficient of resistivity is given for two temperature ranges. (3) Measured at 68°F (20°C). The asterisk denotes Trade Mark Reg. U.S. Patent Office. (Courtesy of Driver-Harris Corporation. From Reference 1.)

low temperature resistance coefficient, high resistivity, high temperature life, and low drift. Materials such as Karma, Advance, and Manganin have temperature coefficients of zero to ± 20 ppm. Other materials such as Nichrome are more suitable for high-temperature applications, but have less favorable temperature coefficients. Resistance stability is a prime system consideration. No matter how good the design of a pot, if the resistance changes with time, the component will cause trouble. Wire drift characteristics are generally not listed in pot specifications, but are made available on request. Wire manufacturers provide good basic guides. The resistance

wire is wound on a support to facilitate its application. Two types of supports are in use—metallic mandrels and insulating cards (Figure 22). Mandrels are normally made of copper, providing good heat conductivity and a reasonably close match with respect to coefficients of expansion. Some type

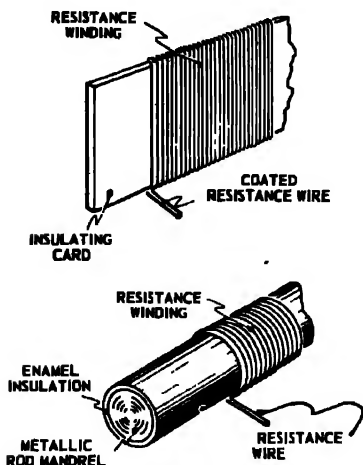


Figure 22. Wirewound resistance elements on card and mandrel supports. (Courtesy Technology Instrument Corporation. From Reference 10.)

of varnish or adhesive holds the winding in position on the mandrel. Other advantages include ease of winding and uniformity. Insulating cards are frequently made of phenolic material and are used when large-diameter resistance wire is required or a flat expanse of wire is called for to permit the use of multiple contact sliders. The heat-conduction properties of cards are poor, and their coefficient of expansion is considerably different from that of the wire wound on it; this leads to strain and slack in the wire under temperature cycling conditions. After the wire is wound around a support, the assembly is formed so that it can be dropped into the housing (Figures 23 and 24). It is held in the housing by some form of mechanical fastener molded in position or by adhesives. One of the problems associated with wound components is maintaining uniform winding tension. If tension is too low, there will be slack in the winding; too much tension stretches or may even break the wire. Wire that is stretched has a smaller than normal cross-sectional area and, consequently, increased resistance. It is easier to maintain winding tension on a cylindrical surface than on a rectangular card, since the corners tend to produce stress-concentration areas.

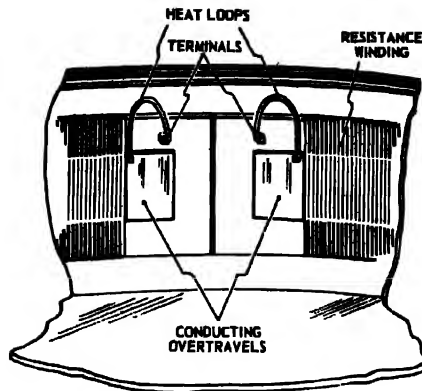


Figure 23. Conducting overtravels at ends of the resistance element. (Courtesy of Technology Instrument Corporation. From Reference 10.)

The slider contact problems in potentiometers are basically the same as those in slip rings. The important difference is that the power levels are considerably lower in potentiometers than in slip rings. Cantilever brushes as well as those supported at both ends are currently in use. A typical unit is shown in Figure 25.

The drive shaft assembly must provide rigid coupling between the input

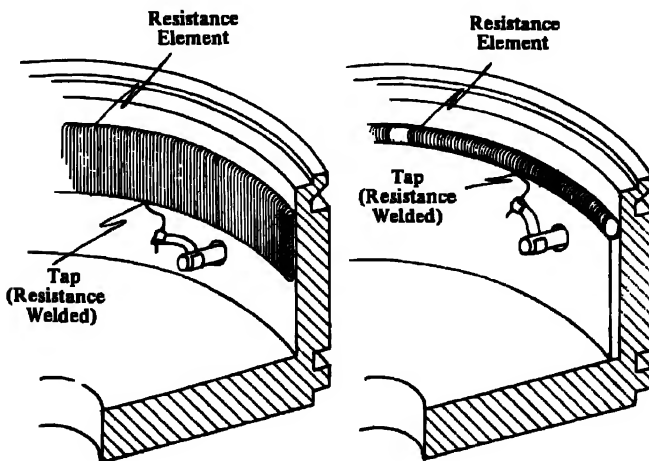


Figure 24. Electrical tap. (Courtesy of Technology Instrument Corporation. From Reference 10.)

312 Slip Rings and Potentiometers

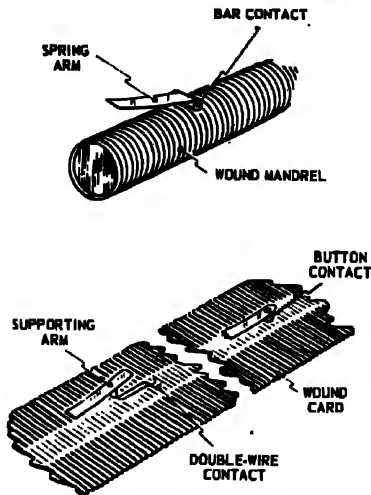


Figure 25. Typical slider contacts for mandrel and card resistance elements. (Courtesy of Technology Instrument Corporation. From Reference 10.)

and the wiper arm without excessive torque levels. It also contains a slip ring assembly that is as important as the wiper assembly (Figures 26 and 27). Our earlier remarks on slip rings are appropriate here as well. The shaft is usually a precision-ground stainless-steel element with precision bearings. One key to the value of the potentiometer is the grade of bearing used. The minimum standard is usually an ABEC-3 quality bearing; ABEC-5

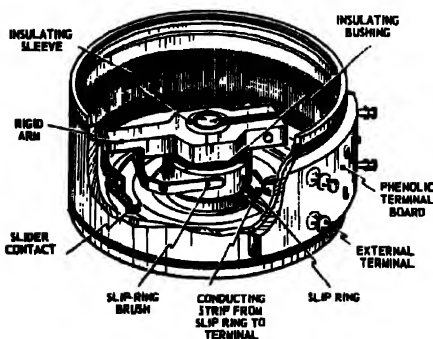


Figure 26. Slider and slip ring assembly forming typical electrical takeoff for potentiometer output. (Courtesy of Technology Instrument Corporation. From Reference 10.)

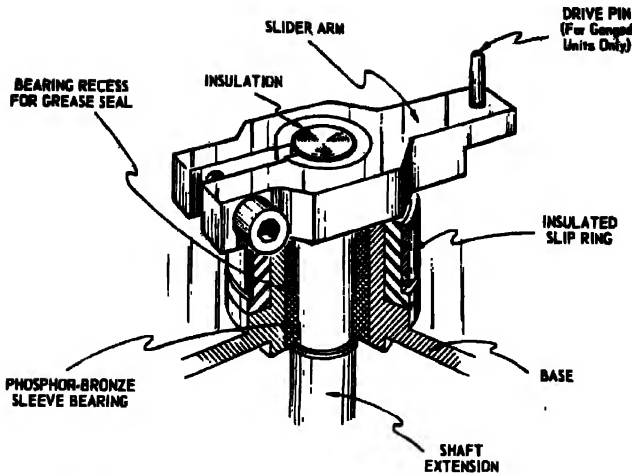


Figure 27. Shaft and sleeve bearing structure of a typical precision potentiometer. (Courtesy of Technology Instrument Corporation. From Reference 10.)

is considered to be high-quality, and ABEC-7 is reserved for absolutely minimal torque applications where price does not matter. To obtain a very quick indication of the type of potentiometer, the systems engineer should first consider the quality of bearings used in the unit.

The terminals or taps on a potentiometer are located at the electrical extremities of the resistance winding and at the wiper. Some designs permit their placement every 90° , every 45° , or some other uniform increment apart. The terminal is connected to the resistance element by a wire that is resistance-welded at both ends. This is a critical area of the pot, since the wire, if not properly supported, may break under shock, vibration, or thermal cycling. A mechanical stop is generally provided at each end of the pot travel so that the wiper travel does not exceed the limits of the end taps.

Covers or enclosures on pots are used for protection against dirt, dust, humidity, and other airborne contaminants. The weak spot is the clearance between the shaft and the supporting enclosure. If O-rings are used, the friction levels increase tremendously. In critical applications, labyrinth seals are used. In applications where considerable heat is developed in the winding, the pot unit should be mounted on a good heat sink; a hole in a metal block is ideal. U-shaped clamps to hold the pot on a metallic surface are satisfactory; servo mount clamps are a poor second. Heat conductive epoxies to couple the case to the heat sink have been used to good advantage.

The construction of a film pot is shown in Figure 28.

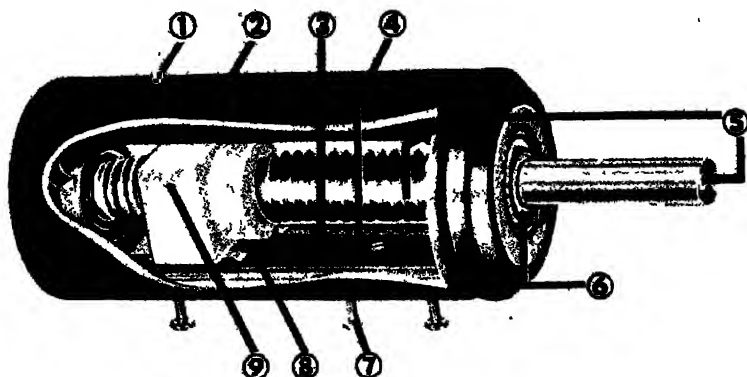


Figure 28. Film-type potentiometer. (1) Optional second resistance element. (2) Aluminum case. (3) Resistance film bonded to high temperature plastic base. (4) Precious metal collector bar inlay, grooved to ensure continuity of wiper. (5) Stainless steel shaft with integral lead screw. (6) Precision ball bearings. (7) Gold-plated terminals. (8) Precious metal, multiple-fingered wiper. (9) Self-lubricating Teflon nut block. (Courtesy of Computer Instrument Corporation.)

6.4.3. Electrical Characteristics

Potentiometer manufacturers have developed more descriptive terminology about the use, design, and testing of their product than the manufacturers of any ten other products that could be named. The terms have been correlated by the Precision Potentiometer Manufacturer Association, and the resulting compilation is available in most technical libraries. Fortunately, only a small percentage of the terms must be mastered by the systems engineer. An excellent digest appears in Appendix B of this book (courtesy of the *Electromechanical Design* magazine).

Potentiometer noise is the first consideration in any component selection. It can be divided into two categories: resolution noise and residual noise. Resolution noise consists of stepwise voltage increments created by the slider moving across discrete turns of the resistance winding. Residual noise consists of sharp random voltage spikes generated as the slider moves across the resistance winding. Since most pot designs emphasize high resolution, or many turns of wire per linear inch of the instrument, resolution noise is highest in these types of instruments. Dirt and contaminants are the prime cause of residual noise. In carbon film pots, the noise is a function of the sliding action between the wiper and the carbon granules in the film. If the

dispersion of carbon is continuous, the resolution noise is low and the prime problem is residual noise. This is a function of surface cleanliness and microfinish. If a film pot is kept clean, the noise should decrease with use because of the polishing effect of the brush on the resistive surface.

The power rating of a potentiometer is a function of the ability of the instrument to get rid of the heat generated. Ratings provided with a pot can be modified by a factor of 2, in either direction, depending on the method of heat sinking or air circulation used. For a given design and usage, a large pot is more likely to survive than a small unit because its mass is greater. A pot housing made of a good conductive material, such as aluminum, is more likely to conduct away heat than a phenolic unit. The ultimate question is whether the heat can be transferred before it damages critical parts of the instrument.

The frequency characteristics of a pot at low frequencies (below 1 KHz) is essentially resistive. As the frequency increases, it must be analyzed as a lumped RLC transmission line. The load impedance becomes very important in determining the response at high frequencies (Figure 29). The effect of

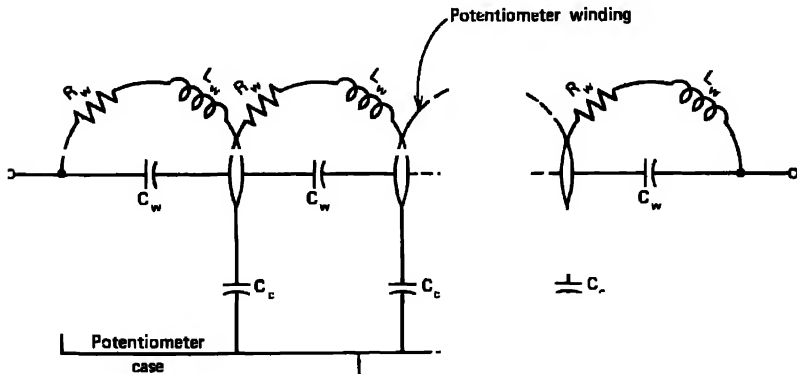
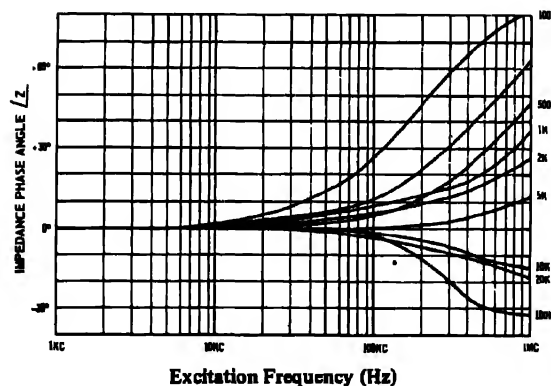
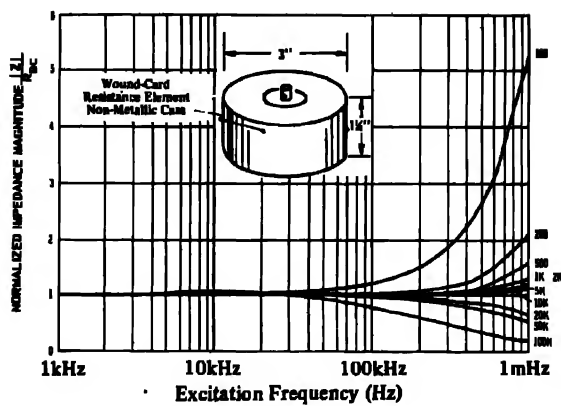


Figure 29. Approximate lumped-parameter transmission-line representation of a wire-wound potentiometer. R_w = Winding resistance per turn; L_w = winding inductance per turn; C_w = interwinding capacitance per turn; C_c = capacitance to case per turn. (Courtesy of Technology Instrument Corporation. From Reference 10.)

the mandrel is to increase further the interwinding capacitance C_w . When a unit with an aluminum case and a unit with a plastic case are compared (Figures 30a and b), it is evident that the impedance matching is more critical when a metal case is used.



(a)

Potentiometer loading, or drawing excessive current through the wiper arm, is a serious problem in precision circuits. The linearity varies as a function of the wiper arm position. Maximum error occurs at about two thirds of a revolution of the input shaft for a single-turn unit. This is an electrical error and is not a function of the composition or construction of the resistance element. It can be minimized by loading the pot with a resistance at least ten times the nominal resistance of the pot (Figure 31). Several other methods are available when this is impractical. Limiting the rotation of the pot shaft to a fraction of a revolution reduces the total error. Placing a resistor in series with the pot accomplishes about the same result. Both methods require increasing the supply voltage if the system scale

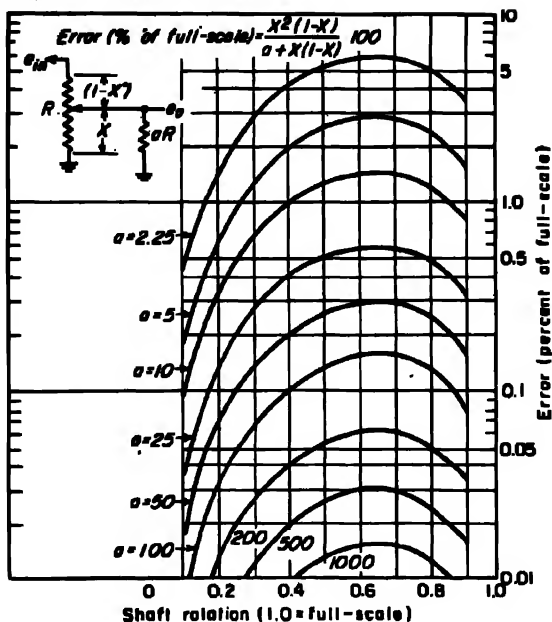


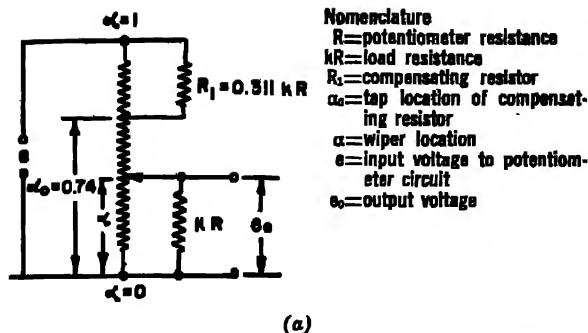
Figure 31. Potentiometer loading showing error produced by load resistors of various values. (Courtesy of Beckman Instrument Inc. From Reference 16.)

connecting a small capacitor from each tap to the closest end point of the pot, the peak quadrature voltage can be reduced by as much as a factor of 5.

6.4.4. Selection

Some of the parameters that should be determined before a pot is ordered are the following:

1. *Function.* Linear, nonlinear, or special function.
2. *Tolerances.* Are the tolerances on linearity or conformity established?
3. *Resolution.* What is the smallest increment of signal required?
4. *Noise.* What is the maximum allowable noise? At what speed is it measured and with what bandwidth?
5. *Special taps.* Are loading or phase angle connections necessary and at what points on the winding?
6. *Power.* Determine maximum power, voltage, and transients that may be encountered. What type of heat sinking is available?



	Tap Location	Shunt Resistor Values	k times Maximum Error
ONE TAP	$\alpha_1 = 0.74$	$k > 10$ $R_1 = 0.311kR$	1.9 percent
TWO TAPS	$\alpha_1 = 0.20$ $\alpha_2 = 0.80$	$k = 20$ $R_1 = 0.66kR$ $R_2 = 0.66kR$	0.5 percent

(b)

Figure 32. Potentiometer load compensation. (a) One-tap compensation. (b) Summary of final pad design parameters. (Courtesy of Benwill Publishing Company. From Reference 19.)

7. *Ambient pressure.* Determine maximum altitude and its effect on dielectric requirements.

8. *Temperature range.* Is the temperature range broad enough to cause problems with the thermal coefficient of resistance?

9. *Electrical frequency.* What is the effect of RLC parameters of the pot at operating frequencies?

10. *Backlash.* Determine maximum allowable backlash.

11. *Torque.* Determine maximum allowable driving torque.

12. *Mounting.* What provisions have been made to mount the unit?

13. *Life.* How many revolutions of the shaft can be anticipated during the use of the end product?

14. *Environmental.* Are there any abnormal conditions of shock, vibration, or humidity present?

15. *Testing.* What will the test method consist of and what will be considered the failure mode?

Characteristic	Wirewound	Conductive Plastic	Cermet	Thick Film
Resistance Tolerance	150 - 250 K ±1 %	500n- 250K ±10 %	100n - 1 meg ±1 %	500n- 2 meg ±10 %
Resolution	$\frac{1}{6,000}$	$\frac{1}{1,000,000}$	$\frac{1}{40,000}$	$\frac{1}{1,000,000}$
Linearity	0.075 %	±0.03 %	0.15 %	0.025 %
Ambient Temperature Range	-55°C +150°C	-65°C +125°C	-65°C +150°C	-55°C +150°C
Power Rating @ 70°C	4.0 watts	2.5 watts	10.0 watts	2.7 watts
Temperature Coefficient	+50 ppm - 0 ppm	+300 ppm -150 ppm	+200 ppm -200 ppm	-400 ppm - 0 ppm
Maximum Continuous Contact Current	200 ma	5 ma	15 ma	5 ma
End Resistance	1/2 resolution	NA	.5 n	NA
End Voltage	1/2 resolution	.01 %	.007 %	.01 %
Jump of Voltage	Resolution	Zero	.05 %	Zero
Tap Widths	Resolution	Zero	Zero	Zero
Tap Current	200 ma	Zero	Zero	Zero
Contact Resistance	<.05 n	1 % RT	2 % RT	1 % RT
Electrical Travel	Variable	Fixed	Fixed	Fixed
Torque	0.5 oz.in.	0.7 oz.in.	1.5 oz.in.	0.3 oz.in.
Useful Life	1,000,000	50,000,000	15,000,000	15,000,000
Rheostat Use	Yes	No	Limited	No
Cost	1	1.4	1.4	3

COMPARISON OF PROBLEM TERMS FOR CONDUCTIVE PLASTIC & WIREWOUND POTENTIOMETERS

These Terms Are Analogous Not Equivalent or Interchangeable

Characteristic	C. P. Term	Wirewound Term (&Units)	Remarks
Allowable End Error	End Voltage (% E in)	End Resistance (Ohms)	Measuring end resistance directly on a C. P. will not give a meaningful reading and may destroy the unit.
Function Travel	Theoretical Electrical Travel (Degrees No Tolerance)	Actual Electrical Travel (Degrees ± Tolerance)	Discrete end points don't exist in a C. P. Specifying a toleranced actual travel is not practical as it can't be measured.
Change in Resistance With Temperature	Resistance Temperature Characteristic(RTC) (% of Total R)	Temperature Coefficient of Resistance (T. C.) (Parts/Million/°C)	C. P. 's do not have a linear T. C. characteristic like wirewounds. RTC is more meaningful and describes the phenomena for C. P. 's.
Spurious Variations In The Output	Output Smoothness (% E in)	Equivalent Noise Resistance (ENR) (Ohms)	ENR is a poor definition for wirewounds. It is disastrous for C. P. 's. Applying an ENR test to C. P. 's will test it as a rheostat and may destroy it.
Steps In The Output	None	Resolution (% E in)	Specifying a resolution value for C. P. 's is meaningless. The term essentially infinite is often used and is not inspectable.

Figure 33. Comparison of various precision potentiometers. (Courtesy of Technology Instrument Corporation. From Reference 10.)

6.4.5. Tolerances

No two manufacturers agree on what the ultimate tolerances should be for a given type of pot. Furthermore, improvements in manufacturing techniques are occurring daily and are changing available hardware. With this in mind, a chart of the tolerances is presented here with the hope that it will hold "true" for at least a few microseconds after this book is published (Figure 33). The chart was compiled by Technology Instrument Corporation, Newbury Park, Calif.

6.4.6. Definitions

A glossary of the terms used in the potentiometer industry appears in Appendix B. It will be helpful when examining competitive brochures and technical papers.

REFERENCES

1. *Mallory Electrical Contacts Handbook*, P. R. Mallory Company, Indianapolis, Ind., 1960.
2. *Slip Ring Assemblies*, Electro-Miniatures Corporation, South Hackensack, N.J., 1969.
3. *Precious Metal Contacts*, J. F. Ney Company, Bloomfield, Conn., 1964.
4. *Raisedlay Electrical Contact Parts*, Englehard Industries, Attleboro, Mass., 1969.
5. *Design Considerations for Miniature Slip Rings and Brush Assemblies*, Poly-Scientific Corporation Blacksburg, Va., 1957.
6. *Applications Manual for Capsule Type Slip Ring Assemblies*, NAS Contract No. NAS 8-5091, Poly-Scientific Division, Litton Industries, Blacksburg, Va., 1961.
7. K. R. North, *Characteristics of Slip Ring Noise*, Electro-Tec Corporation, Blacksburg, Va., 1962.
8. *Precision Sliding Contact Devices Symposium*, November 1965, Poly-Scientific Division, Litton Industries, Blacksburg, Va.
9. *Resistance Shunting of Potentiometer Pickoffs*, Giannini Technical Notes, Autumn 1957, Conrac Corporation, Duarte, Calif.
10. *Potentiometer Handbook*, Technology Instrument Corporation, Newbury Park, Calif., 1968.
11. *Fine Size Resistance Wire for Precision Resistors and Potentiometers*, Driver-Harris Company, Harrison, N.J., 1961.
12. S. Scantsoulis and S. Liss, *Precision Potentiometer Characteristics and Limitations*, *Electrical Manufacturing*, January 1956.
13. *Wirewound and Non-Wirewound Precision Potentiometers*, *Electromechanical Design*, May 1967, Benwill Publishing Company, Brookline, Mass.
14. *Precision Potentiometers*, *Electromechanical Design*, December 1958, Benwill Publishing Company, Brookline, Mass.

322 Slip Rings and Potentiometers

15. *Semi-Precision Potentiometers*, Rubicon Notes #1, Rubicon Corporation, Philadelphia Pa., 1958.
16. *Data File*, Helipot Technical Information Service, Beckman Instrument, Inc., Fullerton Calif., 1958.
17. *Electromechanical Design*, July 1967, Benwill Publishing Company, Brookline, Mass.
18. C. Dawes, *Electrical Engineering*, Vol. 1, 3rd ed., McGraw-Hill Book Company, New York, 1937.
19. *Electromechanical Design*, January 1964, Benwill Publishing Company, Brookline, Mass.

Chapter VII Motor Selection

Motor theory, though well understood for many years, has not stagnated, but has continued to evolve almost without pause for more than 75 years. The advances usually are not breakthroughs in theory but, rather, refinements in technology. They are sometimes so important that AC machines replace DC units (or vice versa); more often, however, the changes reflect manufacturing improvements that yield more foot-pounds per watt for a motor of a given size. The modern-day challenge reflects the twin demands of commercial competition and aerospace reliability. The unprecedented use of small permanent-magnet motors for toys, shavers, and other battery-operated appliances is one example of commercial influence. Linear motors may be the salvation of the present urban transportation system. Aerospace motor developments include the ultimate in precision and smoothness such as "cogless" motors, DC torquers, low-inertia rotors, and resistance to temperature extremes. Eventually all these advances find their way into commercial channels to improve performance and reduce costs. This chapter reviews the basic theory of DC and AC motors, some of the technological advances, and useful peripheral material. The emphasis is on motor selection rather than a detailed study of motor technology.

7.1. DC MOTORS

Three basic types of DC motors are in use today; shunt, series, and compound-wound types. Each designation applies to the position of the field winding relative to the armature (Figure 1). The shunt machine, which also generically includes all permanent-magnet motors, has its field in parallel, or shunt, with respect to the armature.

The power output of a shunt motor can be expressed as follows:

$$\begin{aligned}\text{power output} &= \text{power input} - \text{armature losses} \\ &\quad - \text{field losses} - \text{rotational losses}\end{aligned}$$

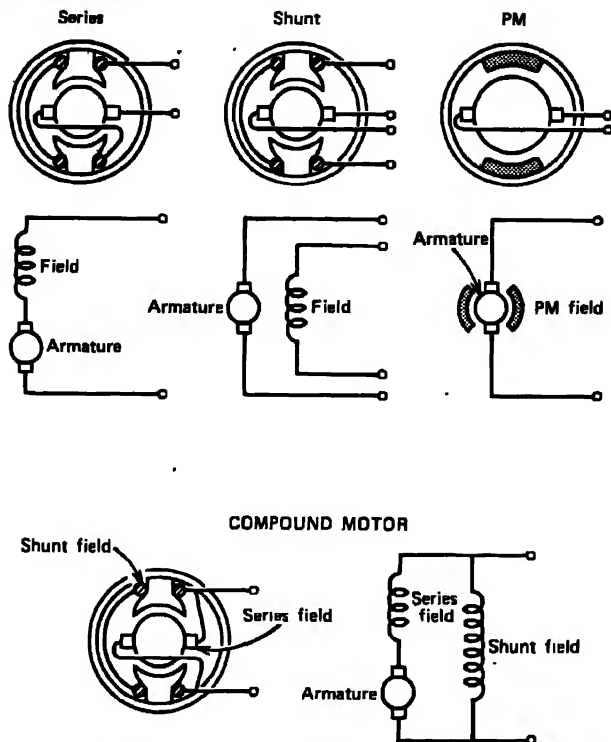


Figure 1. DC motors –schematics of the series, shunt, permanent magnet, and compound motors. (Courtesy of Indiana General Steel Company.)

Rotational losses include friction and windage, hysteresis, eddy current, and commutation losses.

The advantage of having the field winding in parallel with the armature is that the electromagnetic field produced is essentially independent of variations in armature current.

The speed of a DC machine is a function of the following factors:

$$S = \frac{K[V - I_a R_a]}{\phi} \quad (1)$$

where S = speed of the motor

K = proportionality constant

V = terminal voltage

I_a = armature current

R_a = armature resistance

ϕ = flux produced by the armature

It is evident that if the flux is held constant, the speed varies only when there are changes in armature current. Most commercial motors are designed so that the factor $I_a R_a$ is held to about 2 to 5% of V ; consequently the speed regulation is normally held within 5% between no-load and full-load operation. The important advantage of shunt machines is their good speed regulation characteristics.

One of the most important considerations in selecting a motor is obtaining the right torque-speed relationship. For DC machines, the relationship between torque and armature current is as follows:

$$\text{torque} = K\phi I_a \quad (2)$$

For shunt motors the flux, ϕ , is essentially constant; therefore, torque is a linear function of armature current. Since speed is also related to armature current, by the relationship shown in equation 1, the torque varies inversely with speed (Figure 2). This relationship holds until maximum armature

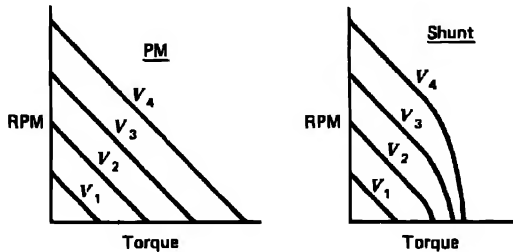


Figure 2. Torque-speed curves for permanent magnet and wound field shunt motors. (Courtesy of Indiana General Steel Company.)

current or maximum torque is reached, where nonlinear characteristics develop. The paramount consideration here is that, in a shunt machine, torque is inversely proportional to speed for a good part of its operating range.

7.2. PM SHUNT MOTORS

The permanent magnet motor retains all the desirable characteristics of shunt motors and also eliminates the field losses. The flux is obtained from



Figure 3. Permanent magnet shunt motor. (Courtesy of Indiana General Steel Compar.

a permanent magnet that produces a uniform field which is independent of voltage, current, speed, or time. This also increases efficiency because, once the motor is paid for, the flux is "free" (equation 1). Since there is no field loss, the motor will also run cooler. The main advantage, however, is lower production costs. Compare the simplicity of the permanent magnet stator shown in Figure 3 with a conventional wound stator. There are no laminations to be stacked, coils to wind, or parts to be machined at assembly. PM motors are also lighter and smaller and generate less noise than wound field motors. The efficiency of a PM motor is usually 15 to 20% greater than a conventional shunt motor. In addition, near-peak efficiency can be maintained over a greater operating range (Figure 4).

7.2.1. Ceramic PM Shunt Motors

Ceramic magnets were developed to meet the need for a low-cost replacement for the conventional permanent magnet material, Alnico V. In addition, they are nonconductors of electricity, lighter than Alnico, composed

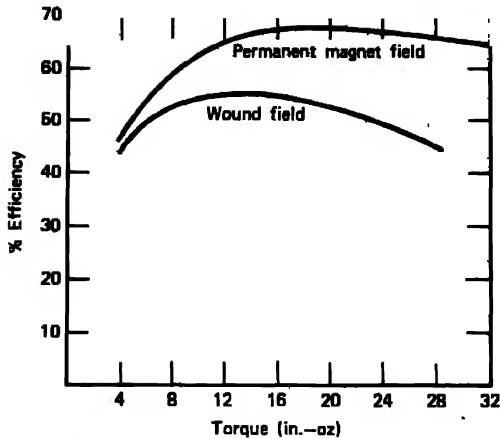


Figure 4. Efficiency versus torque for a permanent magnet and wound field shunt motor. (From Reference 8.)

of noncritical materials, and easy to produce. The trade name for ceramic permanent magnets is Indox. Indox has a very high coercive force that practically eliminates accidental demagnetization. Unlike Alnico, the optimum length-to-diameter ratio of Indox is low, permitting the use of very short magnet lengths.

Typical applications for shunt motors are in medium-starting torque devices requiring good speed regulation, such as fans, blowers, cordless home appliances, and pumps.

7.2.2. Torque Motors

Torque motors are a special form of permanent-magnet shunt motor with a relatively large diameter-to-length ratio. They do not have their own frame, but are designed to be built into the end product (Figure 5). This design produces a very high torque-to-inertia ratio, hence very high acceleration rates. Since the torque is high, no supplementary gear train is required. This eliminates backlash problems and results in an extremely tight coupling to the load. Another advantage is high resolution, limited only by the accuracy of the control equipment. The torque versus current characteristic is also very linear. Torque motors thus make excellent servo motors. Their impact has been so great that many of the very precise servo designs are now DC rather than AC. They do not have their own bearings but depend on the support of bearings used in the end product. This eliminates several parts and lowers cost. The modest size of the device,

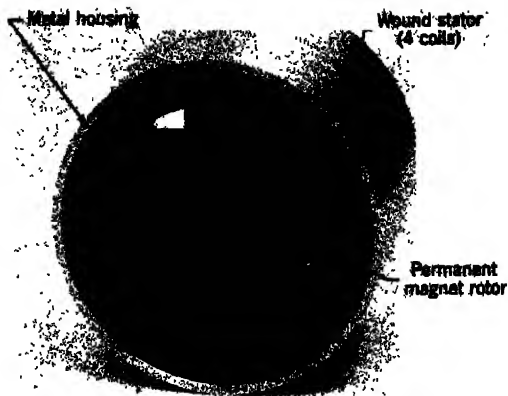


Figure 5. Brushless DC torque motor. (Courtesy of Aeroflex Corporation.)

combined with good reliability, has made this approach very popular in aerospace work.

7.2.3. Brushless Torque Motors

In conventional DC torque motors the stator windings are inserted in slots and create a nonuniform magnetic field. Not only is the field concentrated in relatively small sectors, but the reluctance of the circuit is nonuniform. The result is a nonuniform torque, which may vary as much as $\pm 5\%$, and a nonuniform motion of the rotor; this is sometimes referred to as cogging. A conventional DC torque also has a commutator that tends to produce slightly uneven frictional torque restraints. The slight degree of torque variation is tolerable for most applications, but in aerial camera platforms and similar devices the cogging leads to a degradation of picture resolution. The solution is the brushless torque motor that has a permanent magnet rotor and a toroidally wound field. Since the rotor produces its own flux, no commutator is needed. The windings are uniformly distributed over the entire stator, even though they are electrically isolated. The net result approximates a toroidal winding and leads to a uniform flux pattern. The windings are so arranged that each field exerts a torque on the poles of the rotor in an additive manner. The structure on which the coils are wound is uniform in cross section and produces an even reluctance throughout the magnetic circuit. Brushless torque motors are limited rotation devices.

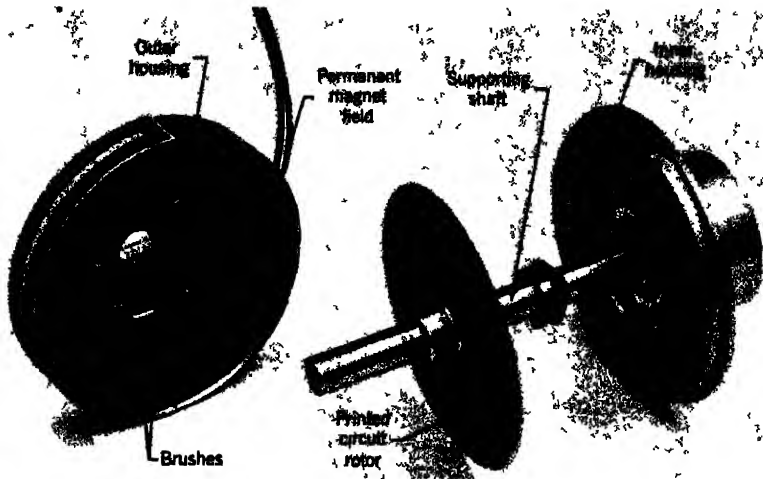


Figure 6. Printed circuit motor. (Courtesy of Photocircuits Corporation.)

Their limits are about $\pm 60^\circ$, but applications normally are restricted to several degrees. These instruments have operated successfully in space as well as submerged in saltwater and a variety of chemicals.

7.2.4. Printed Circuit Motors

The printed (circuit) motor is a DC machine in which the conventional wire windings of a cylindrical armature have been replaced by flat conductors formed on a disk by printed circuit techniques. The motor characteristically has a large, thin envelope. Brushes mounted on the stator of the unit touch the periphery of the conductors on the rotor so that a separate commutator is not needed. The device is an excellent servomotor for the following reasons (Figure 6) :

1. The armature inertia is low, since it contains no iron. Consequently, the response time is low and frequency response is good. Because it responds well to pulse inputs, this device is sometimes used as a stepping motor. The torque-to-inertia ratio is also high.

2. The absence of iron in the rotor eliminates the "cogging" effect found in conventional motors.

Motors with large diameter-to-width ratios are often called pancake motors. These include printed circuit motors as well as conventional motors

whose armature is built from a series of stampings fastened together with squirrel-cage-type conductors.

7.3. SERIES MOTORS

Series motors have their field electrically in series with the armature; consequently flux is directly proportional to armature current. The relationship between torque and armature current is

$$\text{torque} = K'(I_a)^2 \quad (3)$$

where K' = machine constant for a series motor

I_a = armature current

A comparison with equation 2 shows the basic difference between a series motor and a shunt motor; the torque developed by the series motor is significantly greater for a relatively small increase in armature current. The speed regulation is much poorer than in a shunt machine; speed is inversely proportional to armature current:

$$\text{speed} = \frac{K'}{I_a} \quad (4)$$

This equation is a hyperbola; at small values of currents the speed of a series machine can become dangerously high and suitable safeguards must be provided. This is usually some form of starting box. High armature currents produce high torques at low speeds that are ideal for accelerating high inertia loads (Figure 7). Series motors are generally reserved for such

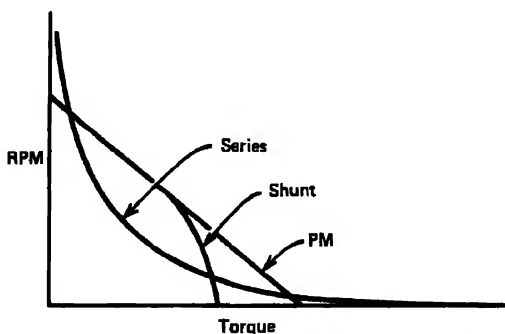


Figure 7. Comparative speed-torque curves for DC motors. (Courtesy of Indiana General Steel Company.)

applications as self-propelled railway cars where very high starting torques are necessary for fast acceleration. Other usages are in cranes, hoists, and elevators.

7.4. COMPOUND MOTORS

Compound motors were designed to obtain the superior torque characteristics of the series motor and the speed stability of shunt machines. Two field coils are provided, one in series and one in parallel with the armature. Some motors have controls that permit the removal of either coil under certain conditions of speed and torque. Motor-generator sets often use compound motors to obtain high starting torque and then short-circuit the series field by a centrifugal switch to get the better speed regulation of a shunt motor. If the series and shunt fields are so connected as to be able to aid each other, the motor is called cumulatively compound; if the fields oppose each other, the motor is differentially compound. The characteristics of the cumulative-compound motor produce greater starting torque than the shunt motor but poorer speed regulation. Unlike the series motor, it has a definite no-load speed and cannot run away if the load is suddenly removed. Differential compound machines produce less starting torque than shunt machines but have speed characteristics that remain substantially unchanged throughout the working range of the motor. It is even possible to have a greater speed at full load than at no load (Figure 8).

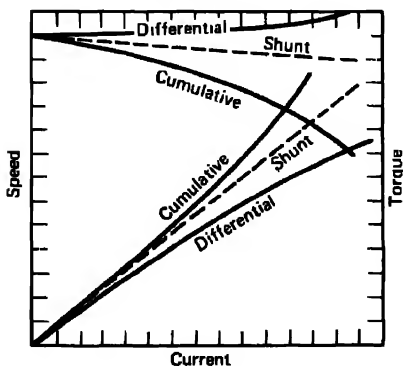


Figure 8. Torque-speed-current characteristics of DC motors. (Courtesy of McGraw-Hill Book Company. From Reference 9.)

7.5. ADVANTAGES OF DC MOTORS

The most important advantage of DC motors is their adaptability to commercial control systems. Today they are being used to control tension in textile processes, velocity of steel mill rollers, synchronization of winding reels, and tension in various processes, including delicate inspection operations. The common denominator is the ability to minutely control DC motor speed over a wide range. Also, many feedback signals in industrial processes lend themselves to DC sensors: photoelectric devices, potentiometers, and temperature and pressure indicators. Recent developments in solid-state DC amplifiers that virtually eliminate drift problems have increased the tendency to use DC control systems for automation processes.

A second inherent advantage of DC motors is their natural adaptability to battery power. Development of rechargeable nickel-cadmium batteries and efficient miniature shunt motors has made the "cordless" appliance industry possible.

The automotive industry is the largest single user of small DC motors; their key criterion is cost.

7.6. AC MOTORS

Three basic groups of AC motors are available today: single-phase, 2-phase, and polyphase. The vast majority of motors sold are single-phase, since they have wider application than any other class of motor. Two-phase motors are used chiefly in precision servo applications and require an appropriate power supply. Polyphase machines are used where large capacities and high efficiencies are required. Minimum size is above 5 hp. A 3-phase power supply requires considerable usage to offset installation costs.

7.6.1. Polyphase Motors

Three-phase motors are used in high power applications where direct current is unavailable or inconvenient. They produce the most horsepower per pound of weight of all AC motors.

Induction and synchronous machines are the two principal types of polyphase motors. Induction motors, in turn, can be either squirrel-cage or wound-rotor units.

The winding in the rotor of a squirrel-cage motor is made of a series of copper bars that look like a small cage. The conductors are interconnected to form a high-current-capacity coil. They are used for heavy-duty constant

speed applications such as lathes and milling machines. They require no commutator and are one of the least expensive 3-phase motors. Torque is proportional to slip; slip is the difference between synchronous speed and actual motor speed. Starting torque is not high.

Wound-rotor induction machines are used when considerable starting torque is required and speed regulation is not a prime factor. The rotor is wound with conventional coils and is equipped with slip rings. The cost is higher than that of squirrel-cage units.

Synchronous motors rotate in exact step with the line frequency. They are used when constant speed is the prime requirement. The rotor is excited with DC power that locks in step with the stator's AC field. Synchronous motors are not self-starting, and an auxiliary squirrel-cage winding in the rotor is used to bring the unit to near-synchronous speed. The main field is then energized, and the unit snaps into synchronous speed. Synchronous motors are considerably costlier than induction motors. Synchronous speed is defined by the following equation:

$$n = \frac{120f}{P} \quad (5)$$

where n = synchronous speed (revolutions per minute)

f = exciting frequency (hertz)

P = number of poles

Polyphase motors are not usually used in accurate systems applications, since they are not readily adaptable to conventional control techniques.

7.6.2. Single-Phase Applications

Single-phase motors are used in such appliances as refrigerators, oil burners, commercial appliances, fans, and machine tools. They are fundamentally constant-speed devices with good speed regulation and starting torque. When speed over a wide range is required, solid-state phase angle controllers must be employed. An important design feature is their low noise level, generally lower than for DC motors. This is a significant factor in the design of home and office products. They are generally less efficient and develop less torque than an equivalent-size DC motor.

Single-phase AC motors are categorized according to the electric technique used to start them. The primary methods are split-phase, capacitor, shaded-pole, and repulsion (Figure 9).

A split-phase motor contains a primary and an auxiliary winding. The auxiliary winding helps to create a field that provides a torque to initiate

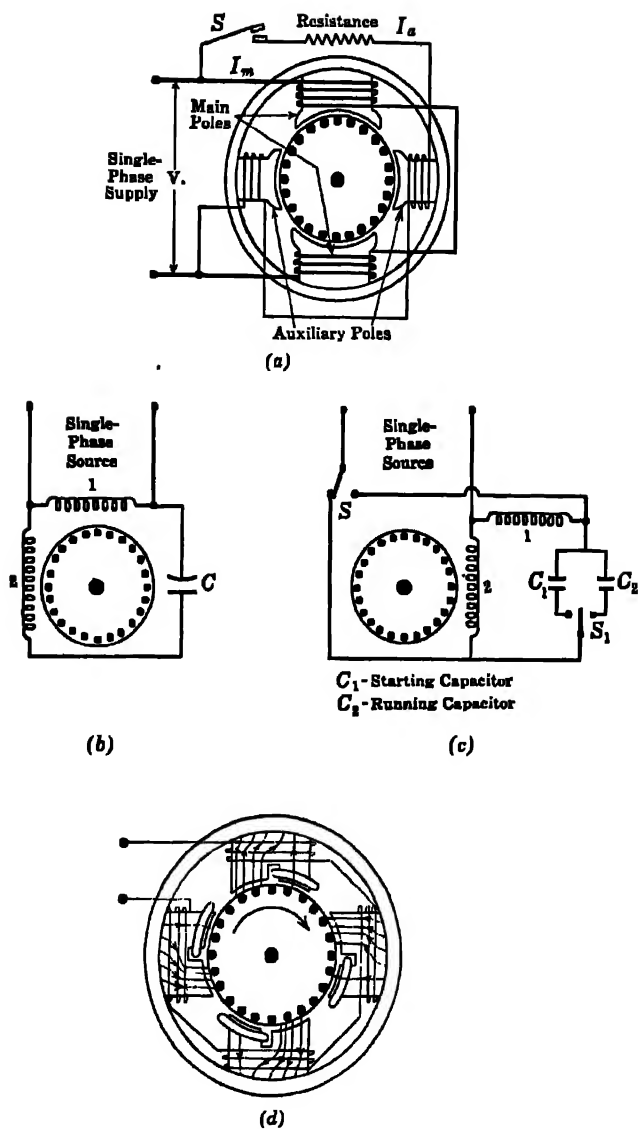


Figure 9. Starting methods for AC motors. (a) Split-phase method of starting single-phase induction motors. (b) Simple capacitor motor. (c) Reversible motor with variable capacitor. (d) Shaded-pole method of starting a motor.

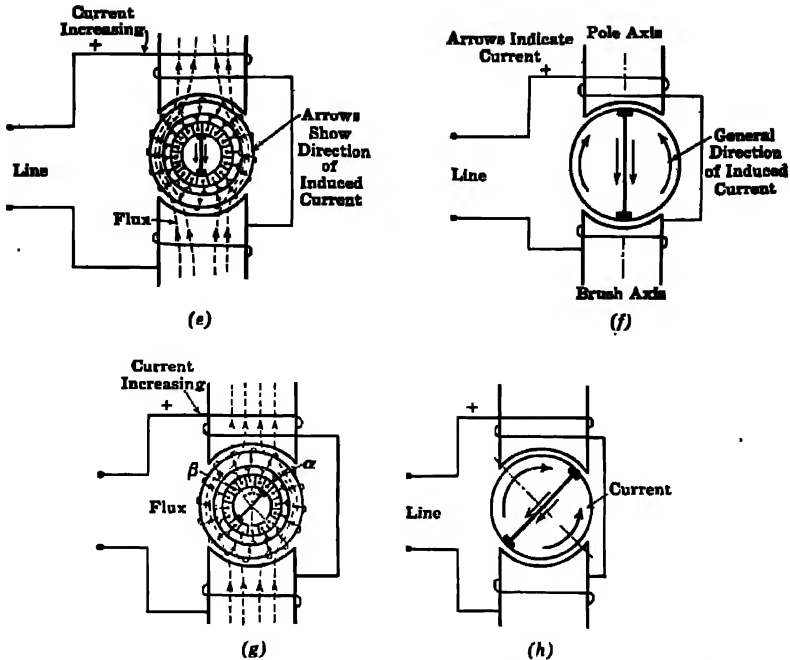


Figure 9. (e-f) Normal position of brushes in a repulsion motor. (g-h) Brush position in repulsion motor which provides starting torque. (Courtesy of McGraw-Hill Book Company. From Reference 3.)

rotor rotation. When a predetermined speed is reached, a centrifugal switch opens a set of contacts that deenergizes the auxiliary winding. The motor continues to accelerate to rated speed in the field provided by the primary winding. This device produces starting torques of 90 to 200% of full value and is considered in the "average" torque category. Speed regulation is good and price is comparatively low. It is not recommended for high inertia loads. The centrifugal switch is sometimes a source of failure.

The capacitor-start motor also contains two windings, but the auxiliary is connected in series with a capacitor. There are three types of capacitor motors: capacitor-start, in which the capacitor is in the circuit only during starting; permanent split capacitor, which has the same capacitor in the circuit for both starting and running; and two-value capacitor motors, in which there are different values of capacitance for starting and running.

Capacitor motors operate more smoothly and quietly than split-phase motors but are more expensive. The capacitor-start motor has high starting torque but contains a centrifugal switch; the capacitor-run motor does not require a switch but has a low starting torque and is the most suitable for light loads such as fans. Capacitor-start and capacitor-run machines produce good starting torque but the capacitor is somewhat expensive; a capacitor switching mechanism is also required. This method is the most efficient of the three, and its speed regulation is good.

Shaded-pole motors are the simplest and cheapest type of AC motor. Instead of an auxiliary winding, they contain a copper loop around a small portion of each pole. This shorted loop, or shading coil, supplies the necessary displacement of the magnetic field to provide a small starting torque. Shaded-pole motors have low torque characteristics, poor speed regulation, and low efficiency. They are typically used in toys, hair dryers, unit heaters, and small timing motors.

Repulsion motors are the AC equivalent of series motors. The unique feature of these machines is a set of brushes on the armature that may be shifted so that the magnetic axes of the armature and field winding are sufficiently displaced to provide a high starting torque. Very high starting torque, average speed regulation, and good efficiency are characteristic of these motors. They can vary speed by shifting the armature brushes. Unfortunately, repulsion motors are expensive and noisy and have brush-maintenance problems. Other types of AC motor have largely replaced them. Typical applications include those requiring high starting torques, such as pumps, compressors, and conveyors.

In every group there is one "maverick"; among motors it is the universal motor, a special type of series motor that operates on AC or DC power. This machine can function many times faster than conventional 60-Hz AC motors. The higher speed makes this motor smaller for a given rating than competitive types. As in DC series motors, starting torque is very high but speed regulation is poor and the speed-torque characteristic approaches a hyperbola. The disadvantage of the motor is limited brush life. Typical applications are vacuum cleaners, electric drills, food mixers, and electric saws.

Hysteresis motors are single-phase timing motors having a rotor fabricated with a hard, magnetic periphery. They start by using the hysteresis losses induced in the hardened steel rotor by the revolving field of the stator. Once the rotor attains synchronous speed, it maintains it by its magnetic retentivity. Starting torque is approximately 100% of rated value; consequently, high inertia loads can be accelerated to rated speed. Common applications are timing devices and gyros.

Servomotors are 2 phase devices designed to produce exceptionally fast

line and the other is supplied from a servoamplifier. The prime criterion for servomotors is the ability to change speed or direction in a very short period of time; the faster the response time the better. To achieve this goal, the torque-to-inertia ratio of the instrument is made very high. The rotor is small in diameter, to minimize inertia so that fast starts, stops, and reversals can be made. The torque-speed curve is very linear. This makes the device compatible with complex servo control requirements. It is also possible to specify damping characteristics as an integral part of the motor performance. These instruments are the most expensive units in the motor spectrum and generally perform in systems where cost and efficiency are secondary to response characteristics.

AC servomotors also can be obtained with several types of damping. This is necessary to control the dynamic response of the associated servo system. Two primary types are available: viscous-damped and inertially damped motors.

Viscous damping produces a retarding torque that is directly proportional to the speed of the motor. The mechanism consists of an aluminum cap mounted on the motor shaft that rotates in a magnetic field. The field is supplied by a permanent magnet. The viscous drag is generated as a result of the eddy currents induced in the rotating cup. Damping is effective in damping out minor oscillations resulting from backlash, end play, and other defects derived from imperfections in the assembly of the motor. The price we pay for these benefits is a lower output torque.

Inertial damped motors consist of an aluminum drag cup rigidly connected to the motor shaft and a magnetic element, mounted on bearings, on the same shaft. During periods of acceleration, the aluminum cup moves with the motor shaft while the high inertia magnetic structure tends to revolve much more slowly. The net result is relative motion between the two rotating elements and viscous drag generated in the aluminum cup. Eventually, the magnetic structure reaches the same speed as the cup and the drag is reduced to zero. Consequently, there is no viscous drag on the motor at rated speed but only when it is changing speed.

7.6.3. Linear Motors

Linear motors are essentially squirrel-cage-type motors that have been produced in linear form (Figure 10). A series of AC coils are stacked side-by-side to form the stator. The rotor is a ferromagnetic rod that moves linearly in much the same manner as a solenoid. When the motor is energized the rod moves linearly at a constant velocity. Unlike a solenoid, the force is constant throughout the stroke. The principal application for these devices is where long strokes are required, since the stroke is limited only by the length of the rod. They are not economically competitive with solenoids

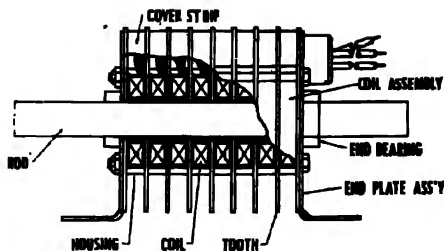


Figure 10. Linear motor construction. (Courtesy of Skinner Precision Industries. From Reference 7.)

until a stroke of several inches is needed. The instrument is reversible by simply reversing any two power leads. Single-, 2- and 3-phase models are currently available. A second version of the linear motor uses a flat armature, with a rectangular motor encompassing it (Figure 11). This type of motor is now being contemplated for use on transportation systems. The motor would be carried in a vehicle suspended over the armature, which would be part of the roadbed. Because its armature need not be in contact with or surrounded by the stator, unique suspension systems can be used. Air cushioning and magnetic suspensions are among those being considered for high-speed transportation systems (Figure 12). It is interesting to note that the first commercial use of the linear motor was by the Kirsch Company for actuating drapery rods.

7.7. MOTOR SELECTION

The most important factors in selecting a motor are torque requirements and ambient environment. Both parameters determine the size and rating of the unit required. Shaft speed is also a prime factor, but is usually not one of the problem areas. For these reasons permanent magnet motor technology is emphasized in this chapter.

7.7.1. Torque

The first step in choosing a motor is to determine the starting torque requirements. There are three basic starting conditions:

1. Starting under no-load conditions with gradually increasing load.
2. Starting under no-load conditions and then suddenly applying the load.
3. Starting under full-load.

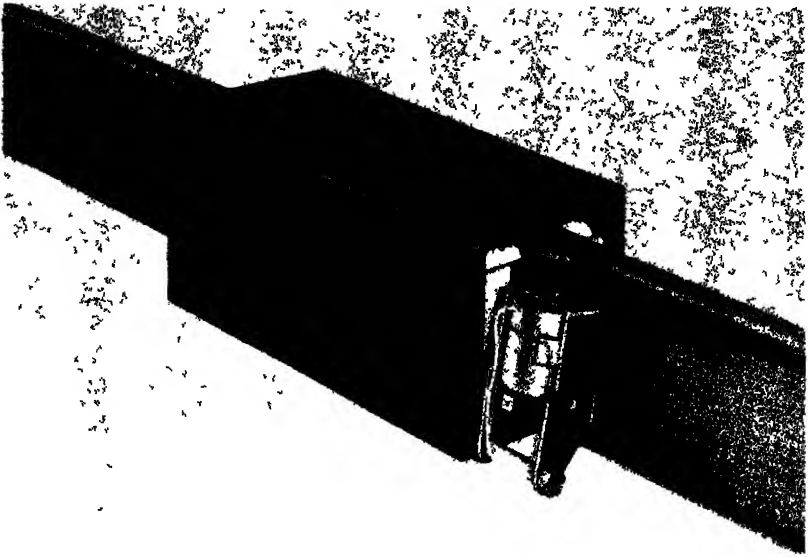


Figure 11. Linear motor mounted on a rail. (Courtesy of Skinner Precision Industries. From Reference 7.)

A good example of a motor starting under no-load is an electric fan. The initial load is composed of the frictional torque associated with the bearings and brushes plus the inertia of the armature and fan blade assembly. As the blade starts to rotate, the horsepower required increases as the cube of speed until the fan is operating at rated conditions.

A hedge clipper is a device that is started under no-load conditions and then suddenly has full load applied. When the knives cut the branches, the motor slows, develops increased torque, and then proceeds to cut. It is important to provide sufficient motor-torque so that the device does not stall when the load is applied. An electric razor has essentially the same loading problem. During the initial no-load phase, any torque available that is not used to overcome frictional loads is used to accelerate the rotating assembly. The governing equation is

$$\text{torque} = I\alpha \quad (6)$$

where I = moment of inertia of the rotor and load assembly

α = angular acceleration

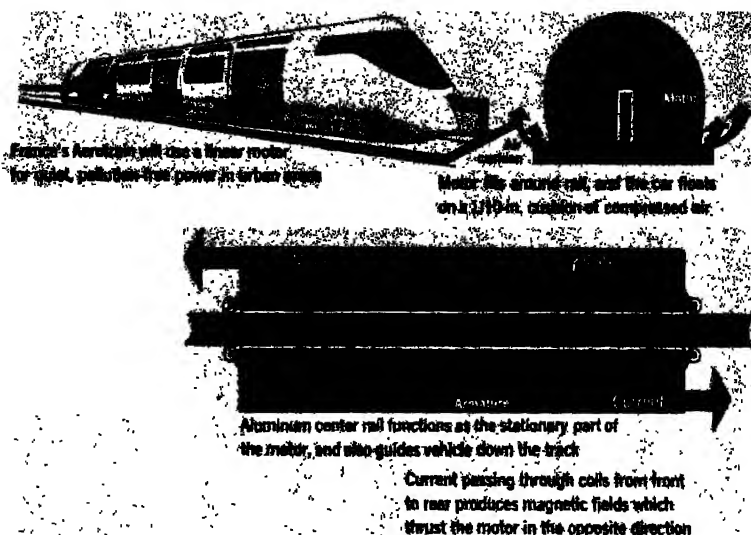


Figure 12. Linear motor used on trains. (Courtesy of McGraw-Hill Book Company. From Reference 10.)

It is necessary to assume that the starting frictional torque is about twice the normal frictional torque due to stiction* effects.

An example of an application where the motor is started under full-load conditions is a windshield wiper. The load is created by the friction of the wiper blade acting on the windshield. If the glass is dry, the starting torque required will be greater than under rain conditions; nevertheless, it is desirable to provide a system that operates at constant speed. This device also requires much faster acceleration than do fans. If a fan is brought up to speed in 3 to 5 sec, it is normally acceptable; this is not good enough for a windshield wiper; hence, higher starting torques and faster response times are provided.

For all three problems it is essential to calculate starting and peak torques, trace the operating range on the torque speed profile, and then determine if the device is operating at optimum efficiency. A very useful characteristic of permanent magnet motors is that the curve of efficiency versus torque is not sharply peaked but is constant between half-load and full load. This

* Stiction is defined as static friction.

makes it possible to "guesstimate" the situation at three-quarter-full load and have about a 25% margin of error in either direction.

7.8. GEARING

A gearmotor is used, for example, to multiply the torque available from a given-size motor envelope or to decrease or increase the output speed.

There are three types of gears in use today; spur gears, helical gears, and worm gears.

Spur gears are efficient, fit into small envelopes, and do not create axial thrust. They are usually stocked in numerous sizes, pitches, and various materials. Helical gears are used when noise must be avoided. The objections to their use are the inherent axial thrust they produce and their increased cost. Worm gearing is used where very high gear reductions and very low noise levels are required. With ratios of about 25–1 the worm gear is self-locking; this prevents the motor from coasting or shifting when deenergized. The big disadvantage of worm gearing is low efficiency—typically 50%, compared to about 90% for spur and helical gears.

When a gear train is selected for a high-inertia load or operation requiring frequent reversals, a worm gear should not be used because the resulting high tooth pressures will cause damage to the train. Spur gears are much more suitable for this type of service. The "backlash" of worm-gear reducers is usually much greater than that of a well-designed spur-gear train. Spring loading applied to worm gear trains further decreases efficiency.

7.8.1. Gear Materials

Both metallic and nonmetallic gears are used extensively and are available from commercial stocks. Nonmetallic gears are commonly used in the first stage of gear trains for quiet operation or when lubrication conditions are marginal. Bronze or steel gears are employed where strength is the prime requirement. Aerospace applications usually require aluminum gears to reduce inertia.

Plastic gears are useful in the following situations:

1. When weight and inertia must be minimized.
2. When the working environment is gritty, abrasive, or corrosive.
3. When the noise level is important.
4. When the application involves flexing of the gearmotor.

Nylon gears are used when vibration damping and low friction are required. Acetals and polycarbonates are creep-resistant, possess low temperature strength, and low moisture absorption. Polycarbonates have good dimensional stability and impact strength. Fabric-filled phenolics are relatively inexpensive and are suitable for stamping operations. Acetals filled with TFE fiber are self-lubricating, creep-resistant, and superior in wear characteristics; they are also the most expensive materials in the group.

7.9. THERMAL CONSIDERATIONS

There are two primary thermal failure modes for motors; increase of resistance in motor windings due to increases in temperature and overheating due to the inability of the device to dissipate the heat generated. The first effect produces lower motor torque since torque is proportional to current; as the armature winding resistance increases, the torque output decreases. In addition, the reluctance of the magnetic circuit increases with higher temperature and the electromagnetic field strength decreases; this, in turn, reduces starting and running torque. Any motor design that is marginal under starting conditions will surely fail under higher ambient temperatures. Overheating causes motor failure due to the breakdown of insulation. The conductors used in the armature of a motor are insulated from one another and the supporting elements by coatings of enamel, yarn, plastics, and other insulators. The entire winding is usually impregnated with an insulating varnish that serves as a binder to provide mechanical and moisture protection. Over a period of time these insulators tend to "age" and lose some of their insulation characteristics at high temperatures. This is due to the formation of microscopic cracks in the insulators that are filled with airborne contaminants with limited conductive properties. When enough of these impurities are embedded in the insulator, a small leakage current may pass between conductors, and the interwinding insulation begins to deteriorate.

To categorize resistance according to these types of failures, the National Electrical Manufacturer Associations (NEMA) established the following classes of insulation:

Class A—105°C maximum winding temperature.

Class B—130°C maximum winding temperature.

Class F—155°C maximum winding temperature.

Class H—180°C maximum winding temperature.

When selecting insulating materials, the NEMA rating should be obtained from the supplier before design work is started. In establishing these categories, it is assumed that the maximum ambient temperature is 40°C

(104°F) and an additional 10°C is provided for motor hot spots; the balance is for the increase in winding temperature due to normal operation. For class A operation, this would result in about a 20% decrease in torque. It is also important to note that for every 10°C increase in winding temperature above the rated temperature the life of the machine is halved.

To illustrate some of the practical thermal problems encountered in motor design, we consider two examples, one mundane and the other "out-of-this-world." They are an automobile windshield-wiper motor and a satellite timing mechanism.

All hot bodies have three ways of dissipating heat: conduction, convection, and radiation. The windshield-wiper motor is normally mounted on the "firewall" at the back of the automobile engine compartment. The majority of the heat generated by the motor is transferred through its housing to the "firewall" by conduction. The ultimate heat sink is the frame of the car, which in turn is cooled by the air flowing past it (convection). The wiper motor is also cooled by the flow of air created by the car fan; it cools the case of the wiper motor by convection. If we want convection to have maximum effect, the case should contain many openings to promote air circulation around the winding hot spots. Practically, this would also allow a prohibitive amount of dirt and fumes to enter the case. The compromise between good sealing and improved convection cooling is a standard problem in all component designs. When cooling by convection is essential, and the environment is dirty, fins may be cast on the case to provide more efficient operation. When the engine is hot, the wiper motor may also be heated by radiation, since the wiper motor normally operates at a far cooler temperature than the engine. The radiant load must be dissipated by normal conduction and convection methods. Thermal problems associated with the wiper motor are rather easy to solve, since the heat sink is available and very adequate; this is not the case with motors used in space environments. The only method of cooling in space is by conduction. The motor must be mounted so that there is a minimum-length thermal path from the "hot spots" to the heat sink. All material in this path must be thermally optimum; copper and aluminum parts are used instead of steel. The real problem is cooling the heat sink itself; it cannot dissipate heat by conduction or convection, but requires radiation techniques. The motor is similarly unable to utilize convection techniques and may receive a radiation load from other hotter bodies. The motor enclosure must be designed to eliminate these radiant loads while placing no impediments in the conductive path. In some applications, part of the space vehicle is hermetically sealed so that 1 atmosphere of gas can be maintained around vital components. A small fan may be used to circulate the gas and cool some of the hot spots by convection. This provides a more uniform temperature within the envelope by

cooling the hotter components at the price of warming some of the cooler ones. Hermetically sealed motors are generally the rule in space work, since severe brush problems develop under vacuum conditions.

7.10. SPEED REGULATION

Two primary methods are used to regulate the speed of permanent magnet motors: series resistance and series-parallel resistance in the armature circuits. For constant field excitation and terminal voltage, increasing the armature resistance causes a change in the slope of the torque-speed curve (Figure 13). This makes it possible to vary the speed from normal values almost to zero. It should be noted that, as the value of resistance is increased, the effect becomes increasingly greater and less stability results. If the value of resistance also changes because of heating efforts, the resulting change in speed will be significant and instability may result.

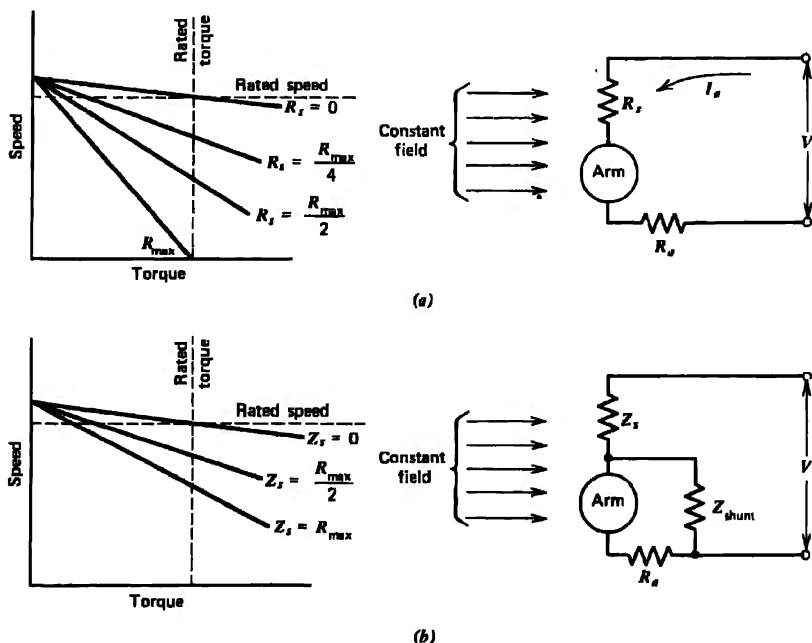


Figure 13. Speed control of a permanent magnet shunt motor. (a) Series resistance in the armature circuit. (b) Series and shunt resistance in the armature circuit.

Example. Calculate the value of series resistance in the armature circuit to obtain zero speed at full load when R_a (armature resistance) is $25\ \Omega$, I_a (armature current) is $0.5\ \text{A}$, R_s is the series resistance to be added to change full load speed to zero, and V (line voltage) is $110\ \text{V}$. At rated starting conditions, then

$$R_s = \frac{V}{I_a} - R_a = \frac{110}{0.5} - 25 = 195\ \Omega \quad (7)$$

Note that this is the maximum value of R ; most applications use only a fraction of this amount to obtain the desired starting conditions.

This calculation will generally result in values about 10 to 15% high, since the effects of friction, windage, and other losses have not been introduced.

The use of series and shunt resistors in the armature circuit provides better speed regulation than the first method, because it is less sensitive to the effects of thermally caused changes. The shunt resistor is also useful in providing dynamic braking at low speeds.

Other techniques for controlling speed involve the conversion of AC to DC power. Silicon diode bridge circuits, silicon-controlled rectifier circuits, thyatron circuits, and saturable reactors are examples. They all produce superior speed control, but require more materials and are much more expensive.

7.11. BATTERY POWER

So far we have considered intentional speed variations due to the nature of the applications. Cordless appliances have a different problem: variation in motor speed due to changes in terminal voltage. The majority of appliances use the nickel-cadmium battery principally because it can be recharged many times. This is usually done simply by connecting it to a common 60-cycle supply. The demodulating and charging circuit is normally packaged in the appliance (Figure 14). The basic nickel-cadmium cell has a nominal terminal voltage of $1.25\ \text{V}$. Normally five or ten are used in series to provide terminal voltages of 6 to $12\ \text{V}$. These batteries are available in button, cylindrical, and rectangular envelopes. The capacities range from 0.5 to $23\ \text{A-hr}$ (Figure 15). The nominal terminal voltage per cell declines from about 1.35 to $0.9\ \text{V}$ before it is "exhausted." This means that the speed and torque of an appliance equipped with a PM field decrease by about 33%. If a wound-field motor is used, the field as well as the armature current decrease, and the speed and torque variation is much larger.

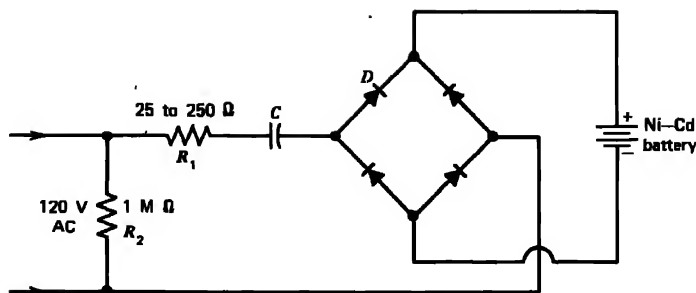


Figure 14. Capacitor-full wave bridge rectification for charging nickel-cadmium batteries (Courtesy of Union Carbide Corporation. From Reference 6.)

Battery ratings are normally based on a discharge period of 10 hours; if current is withdrawn at a faster rate, the capacity of the battery is reduced (Figure 16).

Another rechargeable power source used very extensively in cordless appliances is the alkaline secondary battery. This is not the same as the alkaline primary battery, which is not rechargeable. The newer version was developed to meet the requirements for a power source with a low initial cost. It cannot be recharged as many times as a nickel-cadmium cell, nor does it have the same wide temperature range. Each cell has an average terminal voltage of 1.0 to 1.2 V and functions satisfactorily until it declines to 0.9 V in any 4-hour period. During the early part of its cycle life, there is a very large power reserve amounting to 100 to 200% of the rated ampere-hour capacity of the battery. If it is discharged beyond its rated capacity, total battery life is reduced. However, this reserve power can be used in situations where immediate power is more important than maximum total battery life. During the later part of the life cycle, there is little or no reserve power and the terminal voltage of the battery falls to 0.9 to 1.0 V/cell. Subsequent use produces a chemical reaction that makes further recharging impossible. This characteristic renders it less troublefree than are nickel-cadmium cells.

7.12. BEARINGS

The first step in bearing selection is to decide whether a ball bearing or a journal bearing will be used. The tradeoff is simply performance versus price. Ball bearings are superior to journal bearings in every category except cost and noise. The most serious defect of journal bearings is high

"EVEREADY" BATTERY NUMBER	FORMERLY BATTERY NUMBER	VOLTAGE	NOMINAL CAPACITY (10 HOUR RATE)	CURRENT DISCHARGE (10 HOUR RATE)	CHARGE FOR 14 HOURS AT NOT MORE THAN	CHARGING VOLTAGE	CUTOFF VOLTAGE (10 HOUR RATE)
1.25 VOLTS (cont.)							
RH7	N78	1.25	7 AH	700 ma.	700 ma.	1.35-1.45	1.1
R7.5	N77	1.25	7.5 AH	750 ma.	750 ma.	1.35-1.50	1.1
CH8	N59	1.25	8 AH	800 ma.	800 ma.	1.35-1.45	1.1
CH8T	N59T	1.25	8 AH	800 ma.	800 ma.	1.35-1.45	1.1
R11	N79	1.25	11 AH	1.1 amps.	1.1 amps.	1.35-1.50	1.1
R15	N81	1.25	15 AH	1.5 amps.	1.5 amps.	1.35-1.50	1.1
RH15	N82	1.25	15 AH	1.5 amps.	1.5 amps.	1.35-1.45	1.1
R19	N83	1.25	19 AH	1.9 amps.	1.9 amps.	1.35-1.50	1.1
R23	N85	1.25	23 AH	2.3 amps.	2.3 amps.	1.35-1.50	1.1
6.25 VOLTS							
N67	N67	6.25	900 mah	90 ma.	90 ma.	6.75-7.5	5.5
N70	N70	6.25	1.5 AH	150 ma.	150 ma.	6.75-7.5	5.5
10 VOLTS							
1007	1007	10	4 AH	400 ma.	400 ma.	10.8-11.6	8.8
15 VOLTS							
N63		15	500 mah	50 ma.	50 ma.	16.2-17.4	13.2

Figure 15. Nickel-cadmium battery electrical characteristics. (Courtesy of Union Carbide Corporation. From Reference 6.)

starting torque. Running torque is generally about equal to that of average ball bearings under conditions of full hydrodynamic film lubrication. In practice, this is not always achieved, and journal bearings tend to run two to four times higher than ball bearings. Starting friction is usually about four times as high, necessitating a larger motor than equivalent ball bearing designs. The axial envelope required by journal bearings is much greater than ball bearing designs; conversely, the diametral envelope is significantly smaller. Another major disadvantage is that an oil reservoir and periodic lubrication are required. A journal bearing that is allowed to run dry under

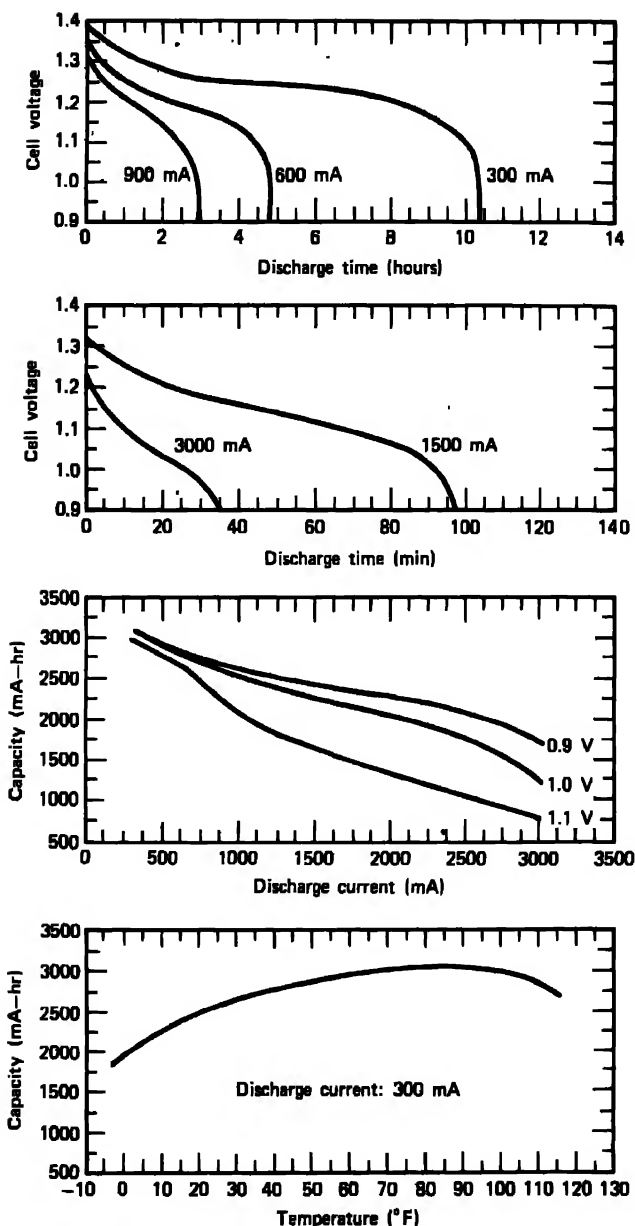


Figure 16. Typical discharge curves for nickel-cadmium batteries at 70°F. (Courtesy of Union Carbide Corporation. From Reference 6.)

heavy loading will result in starting torques 10 to 100 times larger than normal. Solid bronze sleeves impregnated with graphite are used as a semiself-lubricated bearing. The inner diameter is recessed to provide a passageway from the oil reservoir to the inner bearing surface; the presence of graphite helps the formation of an oil film on the bearing surface and also prevents metal-to-metal contact when the motor is stopped. Graphite-type bearings provide good high-temperature characteristics as well. These bearings also resist thrust and radial loads.

A better self-lubricating bearing is the porous bronze type. This design is impregnated with oil, making frequent relubrication unnecessary.

Journal bearings have an almost unlimited shelf life. They do not rust on the shelf, as ball bearings sometimes do. They are also quieter than some ball bearings.

A major variant is hydrodynamic gas bearings that rely on the lubrication and lift of air in journals. Today they are used only in superprecise bearings; future improvements may render them suitable for wider component applications.

7.12.1. Ball Bearing Selection

After determining whether a radial, thrust, or angular contact bearing is required, the quality of the bearing must be specified. Tolerances and fits have been codified into various classes by the Annual Bearing Engineers Committee. These ABEC ratings range from 1, for fair commercial bearings, to 9, for superprecise assemblies used in gyros and accelerometers. Normal commercial bearings are usually rated between 1 and 5. The closer the tolerances on geometry, surface finish, and frictional torques, the better the life of the assembly. Parallelism of raceways can affect friction torque by causing variations in radial clearance and consequent clenching of the balls. Radial runout of the bearing raceways can increase unbalance in rotating masses, such as armatures, resulting in excessive loading at high speeds. The following equation shows the shaking force developed by a rotating member due to bearing runout:

$$F = (2\pi)^2 f^2 M A \quad (8)$$

where F = shaking force developed (pounds)

f = frequency of vibration (revolutions per second)

M = mass of the rotor $\left[\frac{\text{lb}}{386 \text{ in./sec}^2} \right]$

A = one-half the peak-to-peak eccentricity (inches)

The eccentric member is typically the armature assembly. The runout of a bearing is one of the most serious single defects and may result in the generation of tremendous shaking forces and noise. Beside runout on races, single defects on the outer and inner races also contribute shaking forces. A given defect in a race may cause a shaking force every time a ball strikes it and can generate much higher vibration levels than anticipated. The masses of the motor stack, wire, and commutator, as well as the mass of the shaft, act on the bearings under conditions of shock, vibration, and acceleration. The structural design must provide a natural frequency at least 40% higher than the rotational frequency. The spring rate of the shaft must also be high enough to prevent damage to the bearings under dynamic loading. To survive shock and vibration, the bearings must be designed to minimize radial play. Otherwise the mass of the rotor gains kinetic energy relative to the stator and the energy is absorbed on the bearings, causing brinnelling of the balls and raceways.

7.13. LUBRICATION

The major function of a lubricant is to provide a film of oil between rotating and stationary surfaces so that metal-to-metal contact never occurs. Because of its characteristic adhesion and viscosity, oil is dragged along by a rotating shaft or ball and deposits a wedge-shaped film between the shaft and its bearing. The oil wedge is formed as soon as the shaft begins to rotate; the forward motion establishes a pressure in the film that supports the load. This process is called hydrodynamic lubrication, which greatly reduces friction and wear in a bearing (Figure 17).

The second function of a lubricant is to carry away heat generated by friction between rubbing surfaces. Most bearings and races are made of steels such as type 52100 that are not particularly good thermal conductors. If the heat generated by slippage was allowed to concentrate over a small area of contact, overheating and distortion would result. The oil prevents this by distributing the heat to other parts of the bearing, which act as a collective heat sink.

The third function of a lubricant is as a seal against foreign particles entering the bearing—liquids as well as solids present in the environment.

Finally, lubrication protects bearings against corrosion when not in use.

The three principal types of lubricants in use today are grease, oil, and dry-film lubricants. The first two are old standbys; the third is a product of space-age technology.

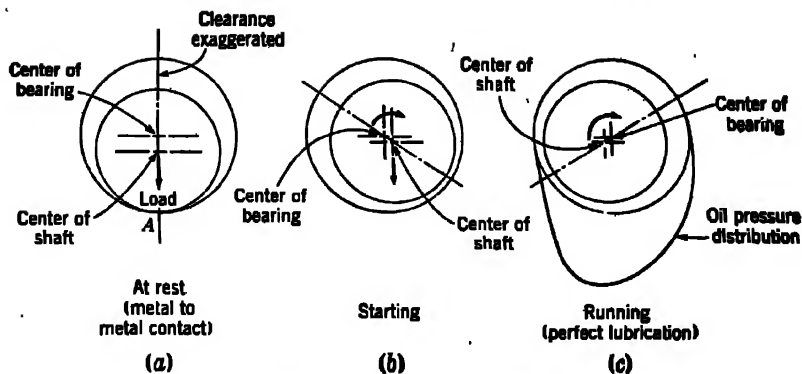


Figure 17. Hydrodynamic lubrication—action of lubricating oil in separating shaft and bearing. (a) At rest—metal-to-metal contact; (b) starting; (c) running—perfect lubrication. (Courtesy of International Textbook Company. From Reference 9.)

7.13.1. Oil Lubrication

The outstanding characteristic of oil lubrication is its relatively low viscosity. Oil can thus easily get into and out of all parts of a bearing. Efficient cooling and cleaning of minute recesses as well as large areas are accomplished. Not only are all areas accessible to oil, but the rates of dirt removal and heat transfer are relatively high. This makes oil the preferred lubricant for most high-speed applications. A second advantage of oil is that it is easy to handle; filling and draining are simple operations. The major disadvantage is the necessity for good seals to keep the oil localized.

7.13.2. Oil Selection

Viscosity is the most important characteristic of oil. It determines the working temperature range, the allowable loading pressure, and the viscous friction of the bearing. When discussing viscosity, two quantitative units are in use: kinematic viscosity and Saybolt universal seconds (SUS).

$$\text{kinematic viscosity} = \frac{\text{absolute viscosity}}{\text{specific gravity}} \quad (9)$$

Absolute viscosity is usually expressed in centipoises, kinematic viscosity in centistokes, and the specific gravity for oil is about 0.9.

To convert from Saybolt universal seconds to kinematic viscosity, the following relationship is used:

$$\text{kinematic viscosity (stokes)} = 0.00226t - \frac{1.95}{t} \quad \text{for } 32 < t < 100 \quad (10)$$

where t = Saybolt reading (seconds),

Water has a rating of about 30 SUS, while SAE #20 motor oil is rated at about 1000 SUS (at 75°F).

Viscosity index is a measure of the change in viscosity over a given temperature range. An oil with a high viscosity index changes viscosity less rapidly with temperature than an oil with a low index. The importance of this characteristic in motor design is that oil with a low viscosity index may greatly increase the viscous friction at low temperatures and prevent the motor from starting. At high temperatures the viscosity may be so low that the oil film breaks down under load (Figure 18). (The high temperature aspect is not normally a serious problem with most small motors.)

Pour point of oil is the lowest temperature at which oil will flow; this should always be considerably below the operating range of the motor.

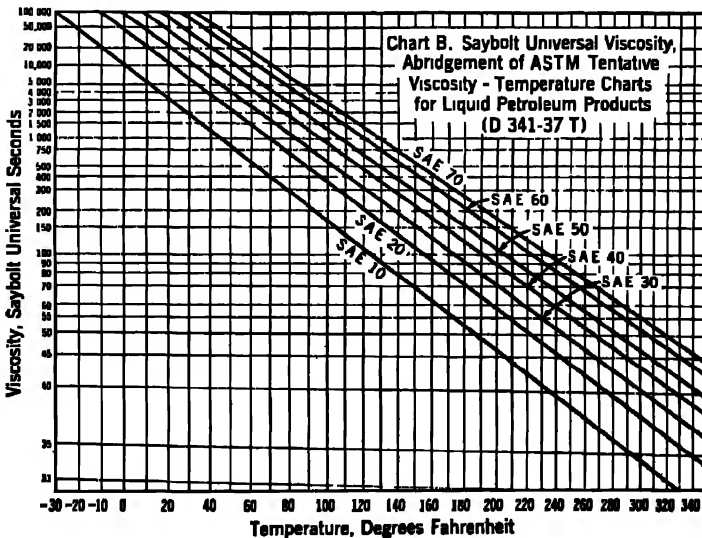


Figure 18. Viscosity-temperature relationship of lubricating oil. (Courtesy of ASTM. From Reference 9).

Flash point is the temperature at which oil gives off ignitable vapors, and fire point is the temperature at which oil will burn when ignited.

Modern synthetic oils have been developed with greatly improved viscosity indices. A typical example is multiduty automobile oils that make starting the car easy on cold mornings, since they are less viscous than conventional oils at low temperatures.

7.13.3. Grease Selection

Grease is essentially a suspension of lubrication oil in a metallic soap base. Its outstanding characteristic is its high viscosity, which makes it easy to retain in motors without elaborate seals. This property also renders it superior to oil in preventing dirt and moisture from getting into the motor. It requires less maintenance as well, since grease does not readily leak out; extended operation without maintenance will not produce catastrophic failures. Grease provides better storage protection against corrosion than does oil. Any soldier who has had to remove cosmoline from a rifle will verify this. The basic limitation of grease lubrication is that it produces a higher starting torque than oil and does not carry away heat as fast as oil does. Therefore, grease is not recommended for high-speed applications.

Grease is rated not on the basis of a viscosity index but of consistency. Consistency is a measure of how easily grease can be squeezed out of the parts being lubricated. It is obtained by dropping a standard metal cone-point downward into a sample of grease. The amount of cone penetration is measured to the nearest tenth of a millimeter. The National Lubricating Grease Institute (NLGI) has set up a scale from 0 to 6, which ranks greases in order of increasing hardness. A grease can be obtained with a low NLGI, combined with a high viscosity oil, or vice versa. The available selection of greases is extremely wide and is constantly expanding.

The choice of a soap base is largely a function of operating temperatures. Calcium soap bases are efficient up to about 150°F. Sodium bases are useful up to 250°F and provide superior sealing characteristics. Lithium and barium bases can be used up to 300°F. Numerous other synthetic oils and bases are available, each one with special characteristics. Unfortunately, the best way of evaluating them is empirically.

7.13.4. Oil and Grease Additives

Various ingredients are commonly added to oils and greases to improve their performance. Oxidation inhibitors prevent the formation of gums and acids by slowing oxidation. Detergents are used to keep insoluble materials in suspension, so that they cannot accumulate on bearing surfaces.

Rust inhibitors are surface agents that protect ferrous surfaces against the formation of rust. Extreme pressure additives provide a low shear-strength film that prevents metal-to-metal contact at high temperatures where the soap base breaks down.

7.13.5. Dry-Film Lubrication

Dry-film lubrication has only fairly recently emerged from the laboratory to be put to limited industrial use. Its purpose is to supply lubrication when every other technique fails. The most important application is for high-temperature, high-vacuum environments; it excels in temperature ranges of 400 to 1000°F and in space environments. Although it was originally developed for the aerospace market, dry-film lubrication is useful for problems involving high-temperature controls for steel mills, ovens, and vacuum process controls.

There are three primary constituents: the lubricant, typically molybdenum-disulfide, tungsten, or graphite; the binder, used to hold the lubricant to the bearing surface; and a solvent, to make the system easy to apply. Dry lubricants can be sprayed, painted, or impregnated into a surface. The Miniature Precision Bearing Company has developed a dry lubricant called Dicronite, which is used in bearings. The process applies a fusion-bonded film of modified tungsten disulfide to a metal substrate in the retainer. As the bearing rotates, the ball rubs off minute quantities of the lubricant until all rubbing surfaces are coated. The lubricant film is about 20 μ in. thick. The modified tungsten disulfide is inert, insoluble and stable up to 900°F in air and 2400°F in a vacuum atmosphere. It has also been used successfully in cryogenic applications. The associated coefficient of friction is 0.025 to 0.090. This technique has been used in conjunction with conventional lubrication in high-speed turbines where the oil pump is driven from the main shaft. If the pump should fail, the dry lubricant would provide emergency lubrication until the turbine could be brought to a stop. The Barden Corporation has perfected a dry lubricant called Bar Temp that is designed to function over a temperature range of -325 to 575°F. The ball separator serves as a reservoir of lubricant in the same manner as in the Dicronite process. Dry lubricants of this type do not contain auxiliary conventional lubricants; consequently, there is no problem of variation in oil viscosity with temperature (Figure 19).

The most important limitations of dry lubrication are relatively high friction, about ten times conventional techniques, and lack of industrial application experience. Life ratings of these products has not been worked out to the same extent as those of conventional lubricants. Another limitation is that solid lubricants do not conduct heat away rapidly or provide sealing

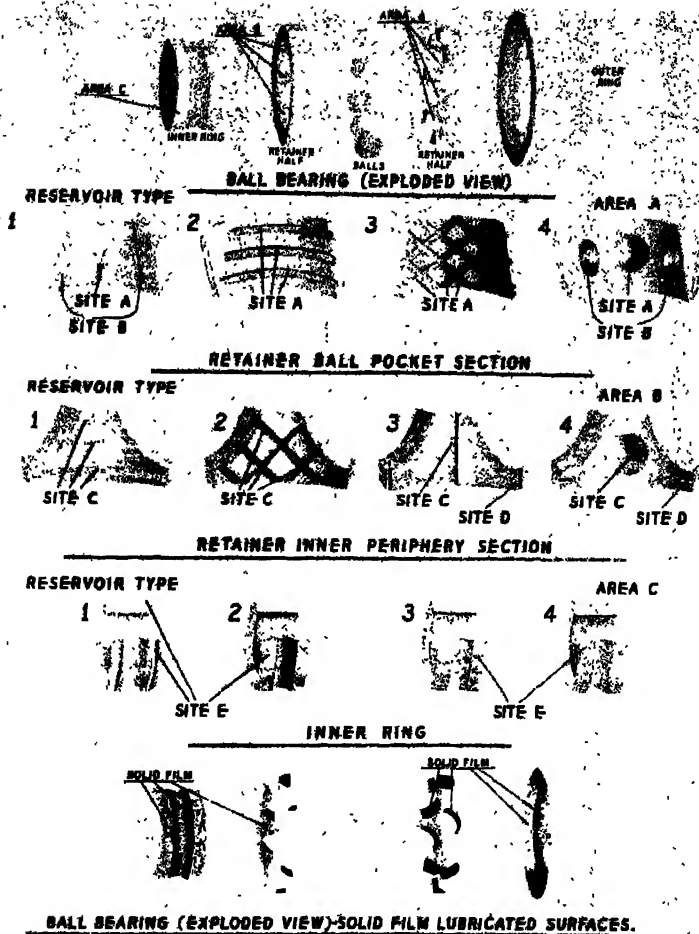


Figure 19. Designs for solid film lubricated ball bearings. Note that "sites" are areas for application of dry film lubricant. (Courtesy of Climax Molybdenum Corporation.)

against contamination. The big attraction of the new method is price. The cost eventually will be very low because only a minute amount of lubricant is used per bearing and application is very simple.

The automotive windshield wiper and the satellite timer previously cited are in very different lubrication categories. If the wiper is designed with sleeve bearings, it should be equipped with self-lubricating bronze bearings

impregnated with oil. This is one part of an automobile that never receives much attention and must take care of itself. If the design utilizes ball bearings, a grease lubricant is required.

The aerospace problem is much more complex. Most conventional lubricants would freeze very shortly. If heat is supplied, the freezing problem is eliminated, but outgassing problems remain. Outgassing is the evolution of gasses from a substance due to low ambient pressure. The process is started when the vapor pressure of the constituent is equal to the ambient pressure. The gas evolved generally will redeposit on any adjacent cold surface, where it may interfere with design functions. Optical parts are particularly affected. In any case, the lubricant is altered by outgassing and does not function efficiently. A hermetically sealed unit is not subject to this problem, but such a unit is not always practical. Some motors designed for these applications are equipped with labyrinth seals around the output shaft so that sufficient pressure is maintained in the motor to prevent outgassing. Dry lubricants are another possible solution but, since the frictional torque levels are high, a larger motor may be necessary and heat sinking problems become more severe. Aerospace applications generally use ball-bearing motors lubricated with light oil and equipped with some form of pressure seal. Heat is often derived from another resistive component in the package or an auxiliary heater.

7.14. ELECTROMAGNETIC INTERFERENCE

Electromagnetic interference is defined as any magnetic or electrical disturbance that prevents the source of this energy from being compatible with other equipment within a given system. Interference originating from numerous sources can trigger other sensitive circuits. Sometimes these circuits are used to detect component or system malfunctions. As a result, malfunctions are indicated where none actually exist. Interference can also trigger auxiliary systems with disastrous results. In a poorly designed system, energizing a radar set may introduce enough interference to trigger a fire-control system. Many electronic circuits are fed with extremely low amplitude signals, which can easily become buried beyond detection by undesirable conducted or radiated random noises introduced into the circuit. One of the major sources of electromagnetic interference (EMI) is motors—both AC and DC.

The major sources of EMI in DC motors are the following:

1. Sliding contact interference at the brushes.
2. Commutation interference.

3. Interference from thermoelectric voltage generation.
4. Thermal agitation provided by current flow through the brushes.
5. Material transfer which takes place when the arc carrying the current between the brush and commutator develops a temperature sufficient to melt microscopic metallic areas of the commutator. The presence of this material, either on the surface of the brush or between the brush and the commutator, causes irregularities in the contact areas and results in an increase in interference.

Only the first two sources are discussed here, since they cause the majority of the EMI.

Sliding contact interference is caused by the shifting of the path between the brush and the commutator. This imperfect electrical path causes random current variations in the form of arcing and sparks.

Sliding-contact interference is minimized by the following methods:

1. Selection of brush material with low contact resistance. A lower brush resistance results in reduced brush temperature, which tends to reduce the level of interference by retarding the rate of oxidation at the contact surface. Reduction in heat loss also improves motor efficiency.

2. Proper rated current for the brush material. As the current density is raised, more heat is generated in the brush contact, speeding the formation of an oxide film on the commutator. This material modulates the current flow and causes higher interference levels.

3. Proper brush pressure. Too low a pressure results in high interference levels due to vibration and chatter; too high a pressure increases interference because of rapid brush wear.

4. High surface finish is achieved by diamond-tipped turning tools; this reduces brush temperatures and increases brush life.

5. Use of individual brush shunts provides lower resistance paths for the current between the brush and the brushholder. This prevents material transfer between the brush and brushholder, which in turn causes heating and erosion because of uneven current distribution.

The second major cause of EMI in a DC motor is commutation interference. This occurs because of the combined effects of rapid polarity reversal of the armature coil current and the collapse of the inductive field of each armature coil. The voltage transients are recurrent, with a repetition rate proportional to motor speed.

Some of the methods used to minimize commutation interference are as follows:

1. Increase the number of coils on the armature and also the number of bars on the commutator to reduce the reactance voltage per coil. For

example, an increase in the number of coils from 5 to 7 decreases the inductance per coil by the square of this inverse ratio:

$$\text{new inductance} = \text{old inductance} \times \frac{(5)^2}{(7)^2}$$

This is about a 50% reduction.

2. Use precision molded commutators to obtain closer tolerances on eccentricity. This reduces brush bounce and arcing at high speeds.

3. Use rigid brushholders to prevent the occurrence of resonances.

4. Use proper brush pressure and current density for the brush design.

7.15. NOISE

Just as EMI may make a motor incompatible with other circuits, noise may make it incompatible with people. Here are some of the ways of minimizing noise:

1. *Dynamic balancing.* All medium- and high-speed motors should be carefully balanced at their maximum rated speeds. This not only decreases noise but also increases the life of bearings and other parts.

2. *Bearings.* Sleeve bearings generally run more quietly than ball bearings. When ball bearings are used, clearances should be as uniform as possible. An ABEC-5 is always a much quieter unit than an ABEC-1.

3. *Brushes.* Brushes with good surface finishes reduce noise as well as EMI.

4. *Thrust washers.* Where thrust washers are used to eliminate end play, nonmetallic ones, such as bakelite or nylon, should be selected to reduce noise.

5. *Gearing.* Worm gear reducers provide quieter operation, if their poor efficiency can be tolerated.

6. *Isolation.* The last step in limiting the noise of motors is to enclose the motor in some type of housing. Sound-absorbent material combined with normal enclosures will help considerably. Isolation of the motor by means of flexible couplings and resilient mountings such as rubber, cork, felt, and springs are also worthwhile.

7.16. REVERSING

One of the advantages of shunt motors is the ease of reversing rotation; the armature leads are simply reversed. Unlike some AC motors, such as split-phase and capacitor-start types, reversal need not be done when the

machine is at rest. Caution must be observed in reversing motors while running, since high currents and torques are generated. This may place limitations on the duty cycle and type of load. High-inertia loads require specially designed rotors and gear trains, if reversing is to be performed as part of routine operation. DC motors characteristically wear one brush faster than the other. This is due to electrolytic action. By reversing the armature connections, this action is equalized between the two brushes, and the average brush life is increased.

7.17. DYNAMIC BRAKING

A motor can be decelerated to stop in three principal ways. The first is by mechanical braking techniques; this requires additional hardware and space, and produces undesirable heat. The second method is called "plugging" and is achieved by reversing the armature polarity until the motor stops and then shutting off the power at the instant of zero velocity. This technique requires a keen eye and lots of practice to prevent the motor from running backward. The third technique, dynamic braking, consists of shorting the armature which causes an opposing current to flow in it, thus creating an opposing flux that produces braking torque. The cautions that were recommended for reversing motors are applicable here. Brush life is often a casualty of sustained dynamic braking.

7.18. OVERLOAD PROTECTION

There is always a possibility that a motor may be damaged or destroyed by unexpected circumstances that are beyond anyone's control. Failure of the motor to start, severe overloading, stalling, blocked ventilation, or unusually high temperatures are the causes. The two ways of sensing danger are by means of fuses and of thermostats. Fuses are designed to sense the heat created by high-armature current. If the motor is overheating because of high ambient temperature or poor ventilating, the fuse will not be effective. Thermostats are sensitive to temperature at one point only. Because of their slow response time, a severe current overload can produce extensive damage before the thermostats sense the resulting heat. To achieve overall motor protection, both fuses and thermostats are required. "Slow-blow" fuses should be selected so that the circuit will not be interrupted every time the motor starts. Thermostats should always be placed as close to the motor hot spot as possible. A thermostat placed near an integral cooling fan will serve no useful function.

7.19. DUTY CYCLE

Motors are generally rated for continuous operation at a prescribed maximum ambient temperature. Under special circumstances, these ratings can be exceeded briefly if the motor is allowed to stop and cool immediately afterward. Another way of achieving this is to supply a fan as an integral part of the motor assembly. Unfortunately, safe overloading of motors on a reduced duty cycle is largely an empirical process. The motor should be equipped with thermostats or thermocouples and bench-tested under simulated, overrated conditions; this is the most practical way of obtaining reliable results for a given configuration.

REFERENCES

1. T. C. Lloyd, *Electric Motors and Their Applications*, John Wiley & Sons, Inc., New York, 1969.
2. J. R. Ireland, *Ceramic Permanent Magnet Motors*, McGraw-Hill Book Company, New York, 1968.
3. C. L. Dawes, *Electrical Engineering*, Vol. I, 3rd ed., Vol. II, 4th ed., McGraw-Hill Book Company, New York, 1937, 1947.
4. Electric Motor Issue, *Machine Design*, December 16, 1965, Penton Publications, Cleveland, Ohio.
5. M. Liwischitz-Garik and C. C. Whipple, *Electric Machinery*, Vol. 2, D. Van Nostrand Company, New York, 1946.
6. *Eveready Battery Applications*, Union Carbide Corporation, New York, 1968.
7. E. J. Milano, *Linear Induction Motors*, Skinner Precision Industries, Inc., New Britain, Conn., 1968.
8. *Appliance Manufacturer*, June 1963, p. 44. Elmhurst, Ill.
9. B. H. Jennings and E. F. Obert, *Internal Combustion Engines*, International Textbook Company, Scranton, Pa., 1944.
10. Going on the Straight and Narrow, *Business Week*, January 20, 1968, McGraw-Hill Book Company, New York.

Chapter VIII Supporting Components

The design of a system usually is a series of compromises between desired component performance and actual performance. Some components are assumed to be difficult to obtain because of unusual performance requirements, while others are simply categorized as catalog items. Very often the exotic components are soon under control because so many people are worrying about them, and the neglected components quietly cause a crisis. In many cases, the catalog items turn out to be anything but "catalog." This chapter examines five mundane components that often are erroneously classified as safe and easily obtainable items: solenoids, brakes and clutches, bearings, power supplies, and operational amplifiers.

8.1. SOLENOIDS

A solenoid consists of a coil for generating an electromagnetic field, a plunger that is acted upon by the field to convert it into useful mechanical work, and an enclosing structure to concentrate the lines of flux generated (Figure 1). The coil produces a field that creates magnetic poles on the plunger. The pole nearer the solenoid is attracted into the coil until it reaches a position where the center of the coil and the center of the solenoid coincide. Most solenoids have a stop in the base of the coil that restricts travel to about half the length of the coil. The motion of the plunger is basically governed by the equations describing the motion of a magnetic pole in a uniform electromagnetic field (References 1 and 2).

8.1.1. Configuration

Two basic configurations are on the market: open-frame and tubular solenoids. Open-frame units have a frame consisting of stamped parts that partially enclose the coil. Tubular models have the coil completely enclosed by metal. The open-frame units are less expensive than tubular units, but,

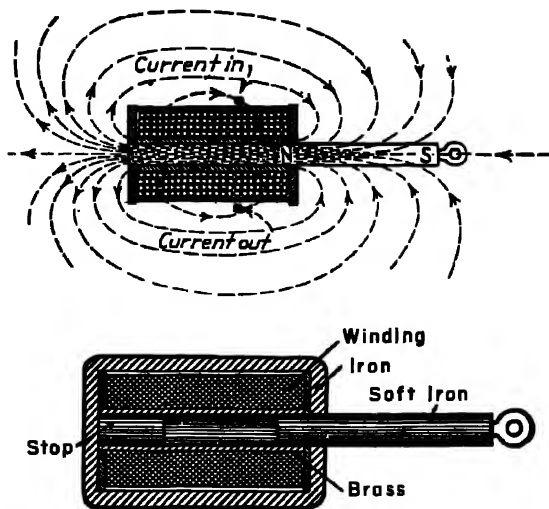


Figure 1. (a) Simple solenoid and plunger. (b) "Ironclad" solenoid and plunger with stop. (Courtesy of McGraw-Hill Book Company. From Reference 1.)

the force produced is smaller. In addition, protection against environmental contamination is distinctly inferior (Figure 2). Open-frame solenoids are commonly used in vending machines, home appliances, and enclosed commercial control systems. Tubular instruments are standard on aerospace work, machinery exposed to the elements, and other systems where the amount of thrust must be a maximum for a given-size instrument.

8.1.2. Construction

The heart of a solenoid is the coil. The design objective is to pack as many ampere-turns as possible into the available space. A detailed design procedure is presented by Roters (Reference 3). From the system designer's viewpoint the most important considerations are the thermal characteristics and structural integrity of the coil. The thermal characteristics are determined by the wire material and the heat sink around the winding. Most manufacturers wind coils using magnet wire rated at 120°C; more accurately, the insulation covering the wire is rated at this value. The insulation is some form of epoxy or polyurethane with a nylon coating. Older solenoids were wound with wire rated at 105°C and covered with enamel, polyimide, polyurethane, and various forms of polyvinyl. By today's standards, these units are marginal and should be avoided. When better than average

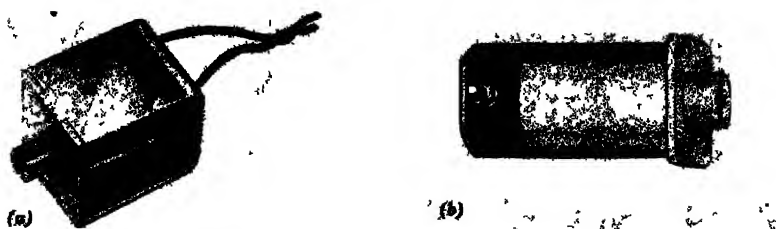


Figure 2. (a) Frame-type solenoid. (b) Tubular-type solenoid. (Courtesy of Detroit Controls)

reliability is required, solenoids are available with 180 and 200°C magnet wire coated with polyester, silicone, teflon, or polyimide insulation. Since most wire manufacturers stock magnet wire with these insulations, the net price increase is seldom significant. The bobbin on which the coil is wound has an important influence on coil life. If the coil is wound on a plastic bobbin, the heat transfer between the coil and the ultimate heat sink is relatively poor. Most of the inexpensive solenoids use molded plastic bobbins because of their low cost. Brass bobbins are much better thermally and are generally specified for high reliability applications. The proximity of the coil assembly to the frame of the solenoid is also important, since the heat must be conducted away from the wire if the solenoid is to survive. The attachment of the lead wires to the coil is a critical process. Unfortunately, this is still an art rather than a science. The only way of determining if this operation is performed correctly is to specify that the solenoid must survive a given pull on the lead wires. Since the wire sizes in various solenoids differ greatly, no standard force tests have been developed. A good rule of thumb is to specify the same value as the tension used in winding the coil. The structural integrity of the coil is further guaranteed by sealing the coil. Some manufacturers "pot" the entire coil or hermetically seal it with a vitreous insulation. Another method called "flashing" consists of passing a high current through the coil for a fraction of a second so that the coatings on the wire fuse together. The less voids left between the wire and other structures the better, since air is a poor conductor of heat; any insulation material is substantially better. Some form of insulator is normally used between the coil and the bobbin. The same thermal criteria should be applied to these components. Class B (130°C) or Class H material (180°C) should be specified.

The shape of the plunger has an important influence on the operating characteristics of the solenoid (Figure 3). The most useful performance curves for a solenoid are the force-stroke curves. Some solenoids provide

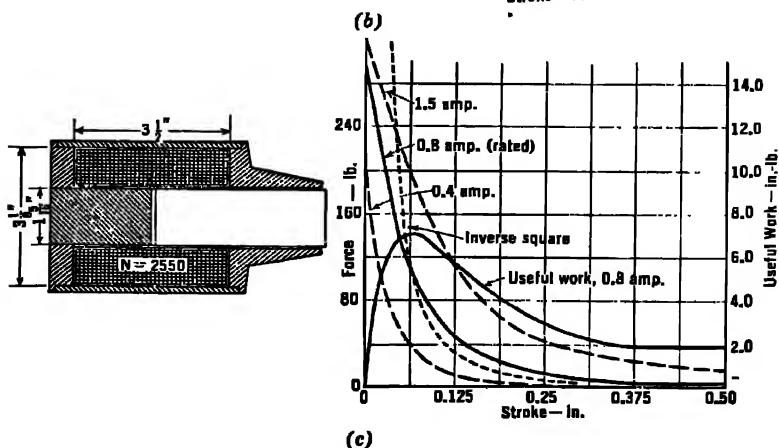
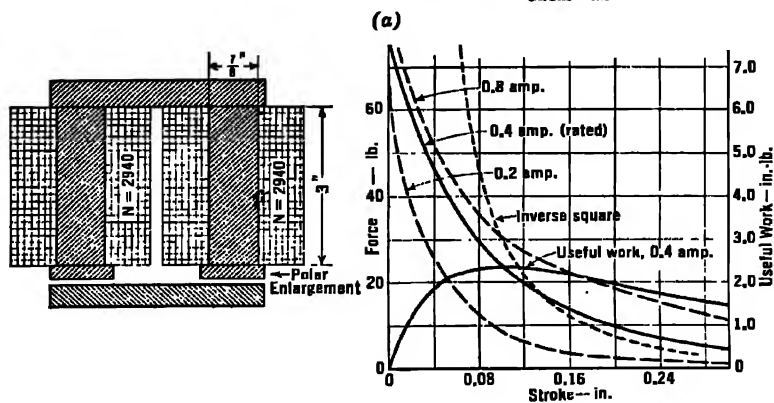
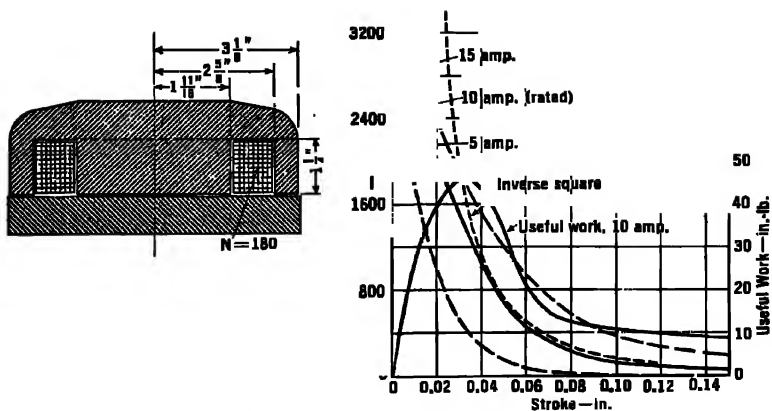
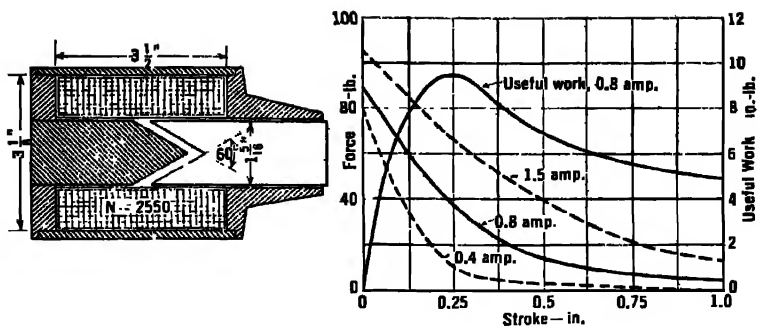
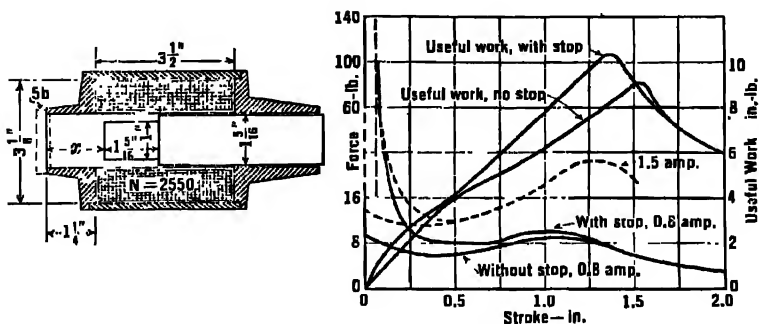


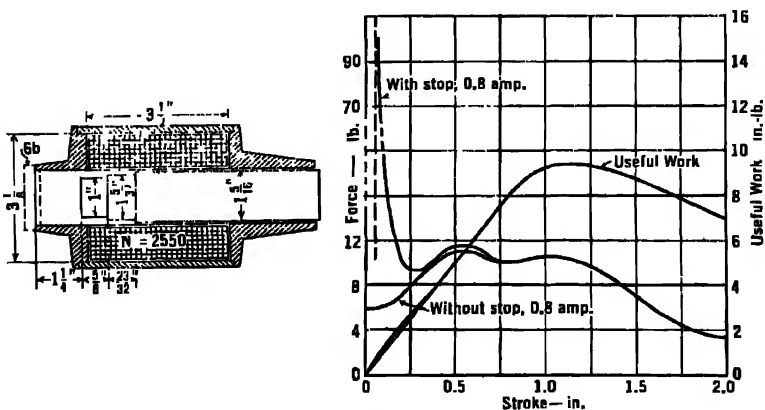
Figure 3. Solenoid configurations. (a) Flat-faced lifting magnet and its force-stroke curves. (b) Horseshoe magnet and its force-stroke curves. (c) Flat-faced plunger magnet and its force-stroke curves.



(d)

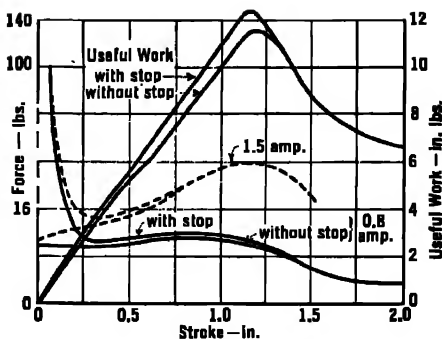
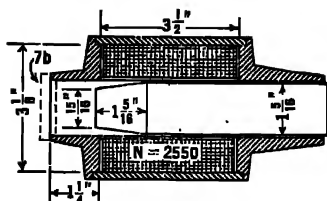


(e)

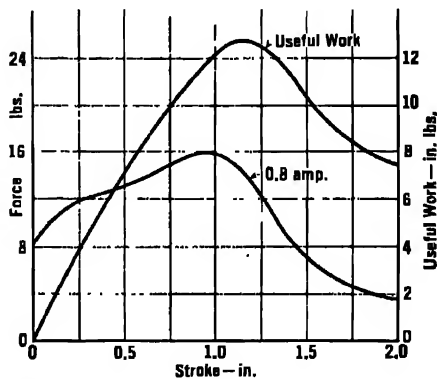
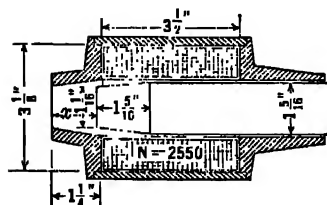


(f)

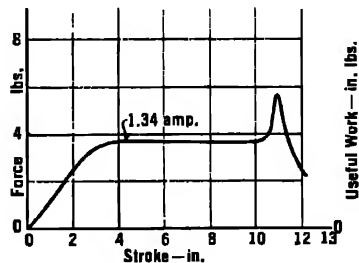
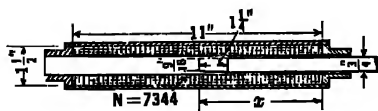
Figure 3. (d) Conical-faced plunger magnet and its force-stroke curves. (e) Cylindrical-faced plunger magnet and its force-stroke curves. (f) Stepped-cylindrical-faced plunger magnet and its force-stroke curves.



(g)



(h)



(i)

Figure 3. (g) Taper plunger magnet and its force-stroke curves. (h) Truncated conical plunger magnet and its force-stroke curves. (i) Ironclad or flux leakage plunger magnet and its force-stroke curves. (From Reference 3.)

optimum performance where a large force is required over a short stroke; others excel at producing a modest force over a long stroke. In each case, the force-stroke curve is determined largely by the shape of the end of the plunger. Seven basic configurations are in use. The key to the proper selection is an index number defined as follows:

$$\text{index number} = \frac{\sqrt{F}}{S} \quad (1)$$

where F = force required (pounds)

S = stroke required (inches)

Once the index number is determined from system considerations, the proper solenoid can be selected from Figure 4. The most commercially available designs are the flat-faced plunger and the 60° conical plunger. Others are available as semispecial units.

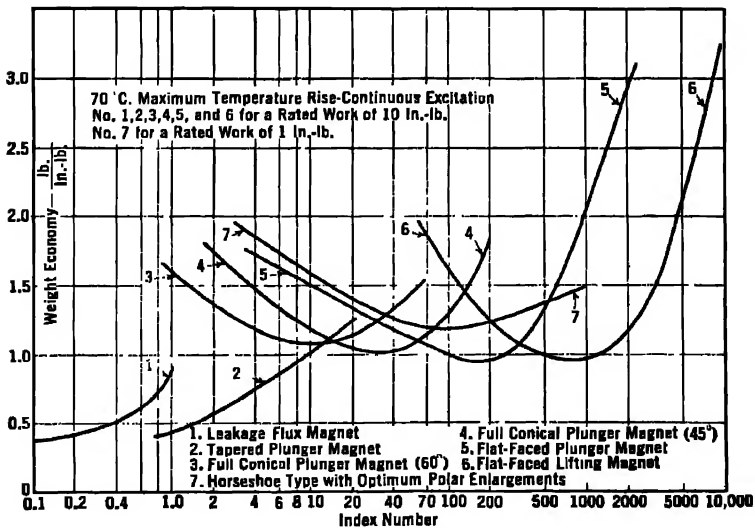


Figure 4. Comparison of weight economies of various types of tractive magnets. (From Reference 3.)

Solenoid plungers are generally fabricated from two types of material: cold-rolled steel and ingot iron. Cold-rolled steel is satisfactory for most applications, but has one deficiency: after a number of cycles the plunger's

magnetic retentivity builds up to the point where it sticks to the stop after the solenoid is deenergized. Normally the plunger is returned to its initial position by spring action, but very often the retentivity is so high that the spring is ineffective. If the spring force is increased sufficiently to overcome this effect, it significantly subtracts from the output force of the solenoid. Ingot iron solves this problem, since its magnetic retentivity is very low. The only problems associated with this material are its greater cost and more difficult machining. Most good solenoids now use ingot iron.

The balance of the solenoid housing also is made of the same metals. The housing should be designed so that it is in close contact with the heat sink; the more surface in contact the better. Standard practice is to use brackets brazed to the housing for mounting pads. A more thermally efficient design is to mount the solenoid in a recess in the heat sink.

8.1.3. Selection of Solenoids

After determining the force and stroke required in the solenoid, the anticipated duty cycle must be established. Duty cycle is defined as the ratio of the "on" time to the total time. Continuous duty solenoids are rated in two ways:

1. Those energized while cold, tested, and left on without exceeding the limits of the design.
2. Those energized while cold, left to stabilize at some final temperature, and then tested while hot.

The amount of force that can be developed by a solenoid, on a continuous basis, is inversely proportional to duty cycle. Most manufacturers have at least two types of solenoids defined as continuous or intermittent. The exact definition of these terms seems to be a question of a lively controversy. In addition to the differences in the definition of continuous duty cited above, intermittent duty may be 1 or 50%. The test conditions for manufacturers' curves are another source of ambiguity. Data accumulated on a 1-oz (of force) solenoid firmly mounted on a Patton tank are less than realistic, as are tests made while suspending the solenoid by a string. Obviously, the systems engineer and the manufacturer must reach a common meeting ground before the printed performance curves assume accurate meanings. A useful solution is a test block approximating the final solenoid mounting configuration. Military practice has been to prescribe a continuous duty test on a specified type of fixture. If the application is at a temperature higher than normal ambient conditions (20 to 25°C), the unit must be derated, since the DC resistance of solenoids increases linearly with temperature (Figure 5).

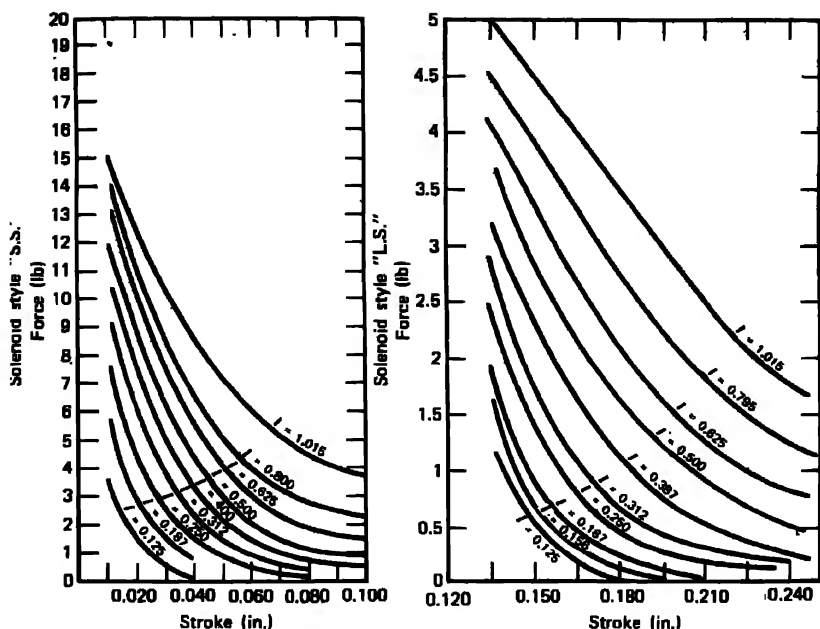


Figure 5. Typical force-stroke and thermal characteristics of a solenoid. Voltage: 18 to 30 V DC; ambient temperature range: -65 to +300°F; weight of solenoid (less plunger) 53 grams (0.117 lb); weight of plunger: 9.0 grams (0.0198 lb).

Coil Number	Duty Cycle Rating (%) ^a	Resistance at 70°F (ohms)	Maximum Permissible "on" Time at 70°F	
			Repeating Cycle ^b	Nonrepeating Cycle ^c
121	100	96.0	Continuous	Continuous
122	80	60.5	Continuous	Continuous
123	50	38.0	60 sec	300 sec
124	25	23.8	40 sec	150 sec
125	10	15.0	25 sec	70 sec
127	1	5.92	6 sec	19 sec

^a Duty cycle is defined as (time "on")/(time "on" + time "off") × 100.

^b Repeat cycle is defined as unit is cycled "on" and "off" more than once, the "on" time not exceeding that shown in the table, at 70 to 100°F ambient temperature. For "repeating" cycle at ambients higher than 100°F consult factory for maximum "on" time and duty cycle recommendation.

^c Nonrepeat cycle is defined as unit energized once, the "on" time not exceeding that shown in the table, at 70 to 100°F ambient temperature. Unit may not be energized for a time interval sufficient to allow it to cool down to ambient temperature. At temperatures higher than 100°F ambient the maximum "on" time for nonrepeating cycles can be conservatively estimated by $t = 0.08 WR (375 - T)$, where t = time (seconds of "on"), W = weight of solenoid and plunger (pounds), T = ambient temperature (°F), and R = resistance at 70°F (ohms).

(Courtesy of Electroid Corporation.)

So far we have discussed parameters that apply to both AC and DC solenoids. There are very distinct differences in the two instruments. The chief advantage of AC units is that AC power is more readily available, particularly for appliances. They are also inherently fast-acting. DC solenoids provide about twice the force produced by AC solenoids for a given pole face area. They are also about half the weight of AC units, since twice as much iron must be provided in an AC solenoid for the same output. Generally, more copper is required in AC units to carry the large reactive power. AC units must be fabricated from laminations to keep eddy current effects within bounds. It is usually difficult to justify the use of AC units in industry. Where only AC power is available, a simple demodulator circuit may be packaged with a DC solenoid. AC solenoids also tend to "hum" during operation.

All the designs discussed so far are pull-type solenoids. Push-type solenoids are also commonly available; the plunger is simply extended through the stop. The most common form is the push-pull type, where one plunger extends through both ends of the solenoid; one end is used for pulling and the other for pushing. By providing an extension through the stop, the pole face area is decreased about 20%; hence the overall force is lower than for comparable pull designs. The cost of both units is about equal; once the tooling is paid for, the differential in material is negligible.

The most formidable problem associated with solenoid selection arises with units that are to be used continuously. It is certainly possible to select a solenoid that is large enough to provide reliable service even when the force produced is greatly reduced by thermal effects; however, this is not always the most economical way of solving the problem. An alternative is to provide a ball and spring detent that will hold the plunger in the actuated position; a cutoff switch deenergizes the solenoid. Some solenoids are available with two concentric coils and a switching device (Figure 6). The first coil, relatively high-powered, is used to actuate the unit. When the plunger reaches its final position, it actuates a switch that energizes the second coil and deenergizes the first coil. The second coil is relatively low-powered, since very little power is required to hold the plunger in the actuated position. The low-powered coil generates much less heat than the primary coil and may be energized continuously without loss of holding power. Another method is to insert, mechanically or electrically, a current-limiting resistor in series with the coil after actuation. Three techniques are used for implementing this method:

1. Mechanical hold-in resistor circuit (Figure 7). When the push button (PB) is closed, full voltage is impressed across the solenoid coil, bypassing the resistor through the NC switch. As the solenoid approaches the end of its stroke, the NC switch opens, inserting the resistor in series with the coil.

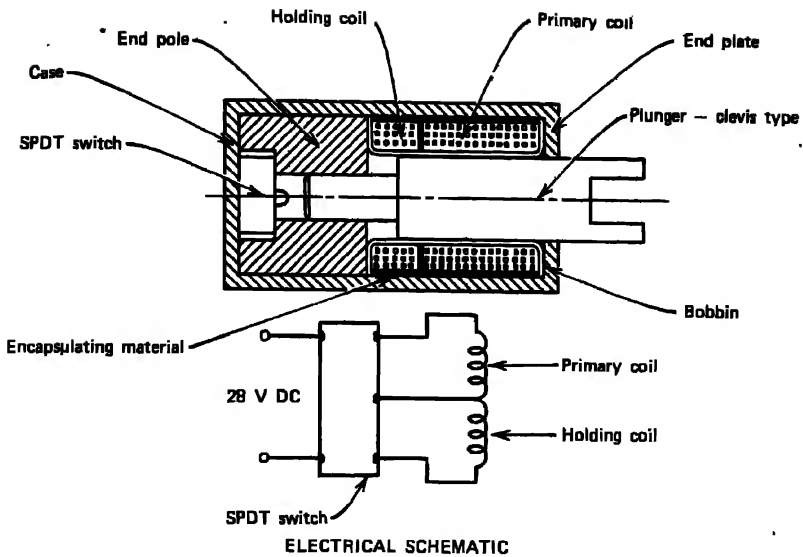


Figure 6. Dual-coil solenoid.

This reduces the solenoid voltage to a point where the power input is high enough to allow the solenoid to hold in and yet stay at a safe power-consumption level.

2. Capacitor hold-in circuit (Figure 7*b*). When the push button is closed, the capacitor discharges through the coil and actuates the solenoid. The unit is then supplied through a rectifier and resistor network at reduced voltage. This is useful only with AC circuits.

3. Transistorized hold-in circuit (Figure 7*c*). When the NO switch is closed, current flows through the transistor to the coil at rated voltage. When the capacitor approaches full charge, the transistor is turned off and the solenoid is powered through a resistor at reduced voltage. The combination of the RC network and transistor is basically a relaxation oscillator with a frequency determined by the RC constant. This constant must be so adjusted that it is long enough to allow the solenoid to actuate.

8.1.4. Auxiliary Circuitry

Successful use of solenoids requires some protective circuitry. The most important consideration is transient protection. When the circuit to an inductive load is interrupted, the reverse voltage generated by the collapsing

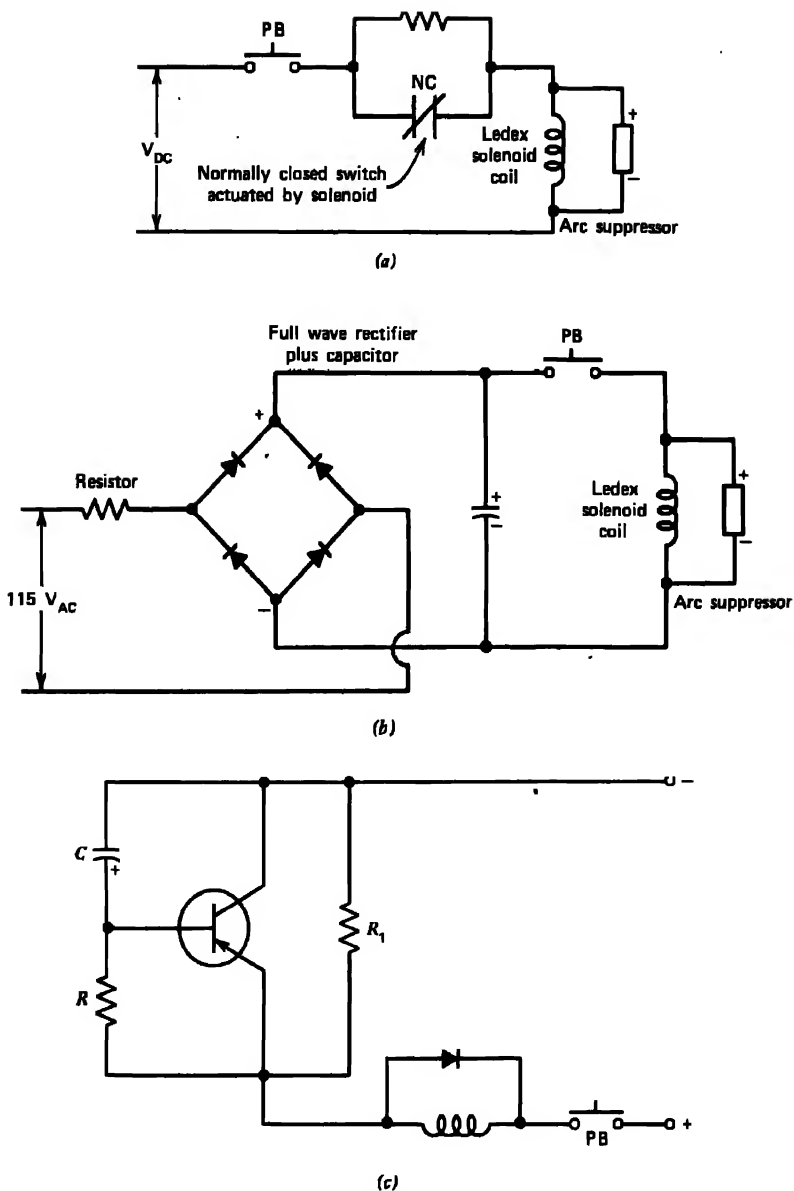


Figure 7. Solenoid "hold-in" circuits. (a) Mechanical hold-in resistor circuit. (b) Capacitor hold-in resistor circuit. (c) Transistorized hold-in circuit. (Courtesy of Ledex Corporation.)

field results in high voltage arcing across the control contacts. It also subjects the coil to voltages many times the nominal rated value. The standard solution is to connect a diode-capacitor network in parallel with the solenoid coil. The only deficiency in this network is that it slows the solenoid response time. When the highest operation speeds are needed, a diode-capacitor-Zener diode network is used. Unfortunately, it does not have so high a peak reverse voltage rating as the diode-capacitor network, but it is adequate for most applications.

Modern solenoid usage has gravitated toward pulse networks for energising solenoids rather than direct current. By this method units rated for several ounces of thrust can be made to produce a pound of force. The normal force multiplication ranges between factors 5-10. The technique consists of impressing a very high voltage on the coil for a very brief duration. In this way the flux density is much higher than normal and the thrust is greatly increased. If this power input were used for any appreciable amount of time, the unit would burn up. The power is applied just a bit longer than the response time of the unit and is then cut off. Thus the heat generated can be easily dissipated to the heat sink. Typical pulse durations range from 10 to 40 msec. Voltages as high as 10 to 15 times normal are used. The limiting factors on this technique are magnetic saturation of the solenoid structure and the insulation resistance of the coil. The technique is used extensively in computers, printers, and other high-speed equipment. The pulsing network is simply an RC network with some form of solid-state switch for applying the energy to the solenoid. The circuit is commercially available and is sometimes called a "pulser." Like all good things, it is good as long as it is used in moderation.

Another useful circuit, commercially available, is the full-wave rectifier. It may be obtained in solid-state form or as a selenium rectifier. It is used for converting AC power to DC for use with DC solenoids.

8.1.5. Rotary Solenoids

Rotary solenoids have the same basic construction as linear solenoids, except for the rotating plunger (Figure 8). The plunger, or armature assembly, is supported by three balls in inclined races. The races are formed in the solenoid housing and the armature. When the unit is energized, the armature pulls in but, because of the motion of the balls in their races, it also rotates. Rotation continues until the balls have traveled to the deep ends of the races. The ball races are formed so that the initial slope is steep; this results in maximum torque at the beginning of the stroke. The slope decreases as the axial pull increases; therefore the net torque of the solenoid is relatively constant for its entire travel. In addition to the primary rotary

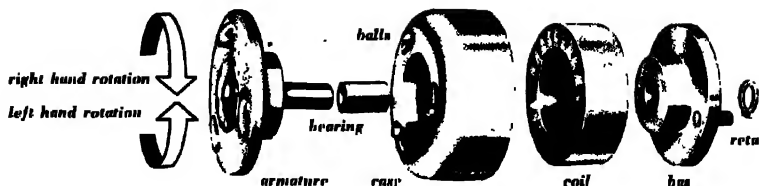


Figure 8. Rotary solenoid mechanism. (Courtesy of Ledex Corporation.)

motion, there is also an axial stroke of the solenoid. The normal range of rotation is up to 95° . Torque is inversely proportional to the total length of rotary stroke. For example, a unit with a torque of 5 lb.-in. and a 50° stroke, will develop 10 lb.-in. if the stroke is reduced to 25° . The Ledex Corporation designs its solenoids to deliver a relatively flat output torque at a 25% duty cycle. In highly intermittent use, such as 10% duty cycle, power and magnetic saturation are increased. This results in a higher starting torque, but a faster reduction in torque, as the armature progresses through the rotary stroke. Since most loads have some inertia, the effect of less torque toward the end of the stroke is usually negligible. In continuous duty, where the magnetic saturation is lower, torque output increases slightly toward the end of the stroke.

8.2. CLUTCHES AND BRAKES

The classical approach to clutches and brakes deals with each as a separate entity. As a result of "natural" selection, the clutches and brakes most used today operate on common principles. Since they are so similar, brakes can be considered to be a special form of clutch that has one plate rigidly fixed by an auxiliary device. This section deals briefly with mechanical devices and at greater length with electric instruments, since they are more commonly used in modern system work.

8.2.1. Positive Clutches

Mechanical clutches are divided into two broad categories: positive and friction clutches. Positive clutches function by metal jaws, or projections on the driver shaft, engaging corresponding parts on the driven shaft. Friction clutches operate by pressing one driving surface against a corresponding driven surface. The basic difference between the two types is that positive clutches will break but will not slip, while friction clutches will slip but will

not break. Positive clutches date back to the dawn of the industrial revolution when they were used in waterwheel transmissions and windmills. They are intended for applications where shafts are meshed at relatively low speeds. Modern applications include shaft coupling where absolute synchronization is important. Since these devices do not slip, they do not develop any heat, and wear is minimal. Size is small compared to other types of clutches, and cost is low. Two examples of positive clutches are shown in Figure 9. The square-jaw clutch is the least expensive but is only suitable

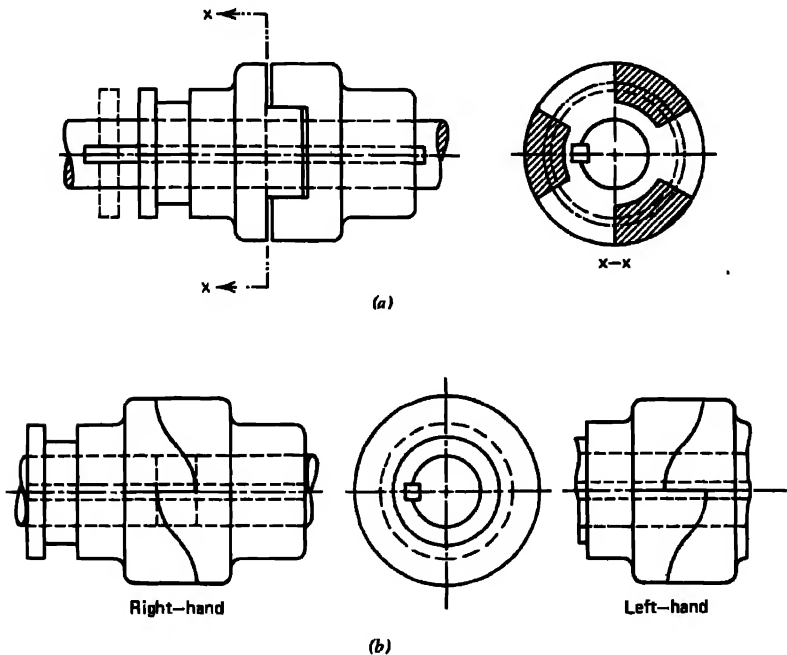


Figure 9. Positive jaw clutches. (a) Square-jaw clutch. (b) Spiral-jaw clutch. (Courtesy of International Textbook Company. From Reference 5.)

for use at very low speeds, typically under 10 rpm. At the moment of engagement, the shock to both shafts is quite high. They are sometimes used in presses where the transmission is automatically decoupled after completion of the working cycle.

The spiral clutch can be used at higher speeds than the square-jaw clutch, but it drives in only one direction. Engagement at speeds up to

150 rpm is possible with low-inertia loads. The shock load to the system is quite high.

Multiple-tooth clutches, containing many interacting teeth, can be engaged at moderately high speeds—typically up to 300 rpm. Engagement is smoother than that found in jaw clutches, but wear on teeth is higher. Some units are available with electrical drives. Initial slip is greater than that found in other positive clutches.

8.2.2. Friction Clutches

Friction clutches were developed to meet the requirements of modern industry and transportation. They can successfully engage power trains at substantial loads and high speeds. The slipping characteristic enables the device to absorb most of the initial shock of engagement without damage. The load is then brought up to the required speed with minimal shock and jerk. Many applications use a friction clutch as a safety device; by slipping when the torque transmitted through it exceeds a safe value, it prevents the breakage of parts in the transmission train.

There are two principal types of friction clutches—axial and rim clutches. Axial clutches have the contact pressure applied in a direction parallel to the axis of rotation. Axial clutches are subdivided into cone and disk clutches. In rim clutches the contact pressure is applied normal to the rim. They may also be made into band and block clutches. An example of a block-type rim clutch is shown in Figure 10. It is used to connect two shafts in line. The block of one shaft engages the rim of the second shaft when the clutch is engaged. It is used in applications where a moderate degree of slip is permissible and duty cycles are low. It is not considered applicable

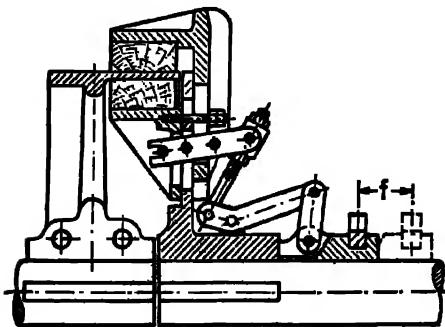


Figure 10. Rim clutch coupling. (Courtesy of International Textbook Company. From Reference 5.)

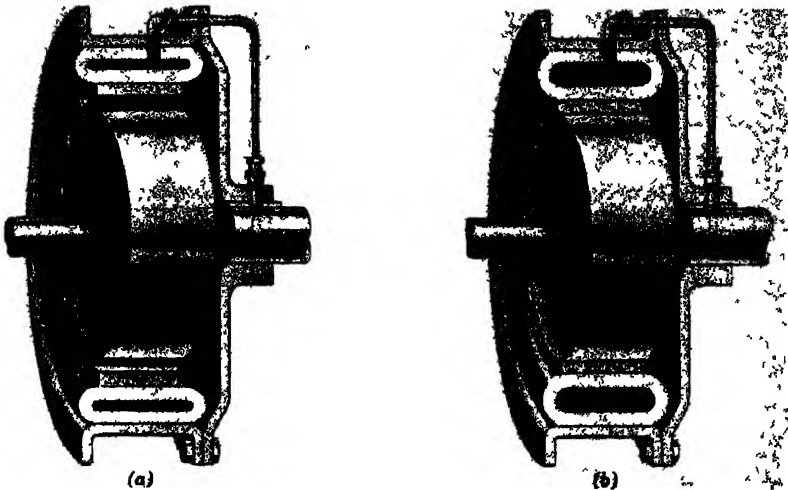


Figure 11. Pneumatically operated rim clutch. (a) Disengaged. (b) Coupled to the outer rim. (Courtesy of Fawick Airflex Corporation.)

in modern machinery. The band clutch shown in Figure 11 is a modern device that uses bands or segments of bands to couple two concentric shafts. It uses compressed air to expand a circumferential tube that in turn presses a series of friction surfaces, or bands, into contact with the rotating drum. This design offers the following advantages:

1. Uniformly applied pressure, resulting in even wear and long life.
2. Torque that is applied at a maximum diameter, providing high ratings per unit volume of the device.
3. Rotation of all slipping surfaces at uniform velocity, contributing to long life.
4. Good heat-sinking capabilities.
5. Modulation of the air pressure, making partial slippage possible for special applications.

A typical cone clutch is shown in Figure 12. With a relatively high capacity combined with low cost, it is best used in applications where engagements occur at low to medium speeds, since the engagement of the mating cone surfaces normally produces some degree of jerk. It is commonly used in punch presses, drill presses, and other low-power machine tools. The most commonly used mechanical friction clutch today is the disk clutch.

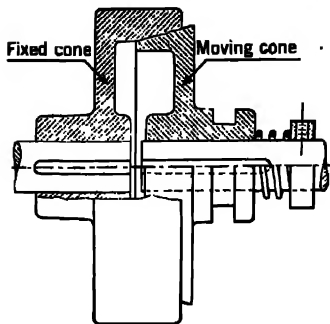


Figure 12. Cone clutch. (Courtesy of International Textbook Company. From Reference 5.)

(Figure 13). Its basic advantage is the employment of multiple coupling surfaces within a given volume. This provides exceptionally high rating per unit volume of the clutch. The torque capability is directly proportional to the number of plates engaged. Some designs use as many as 50 pairs of plates. Most units run dry with conventional asbestos or bronze friction surfaces, while others are designed to run in an oil bath and utilize steel disks. The bath provides better thermal characteristics than dry units. Another advantage of multiple disk units is that failure of one pair of plates

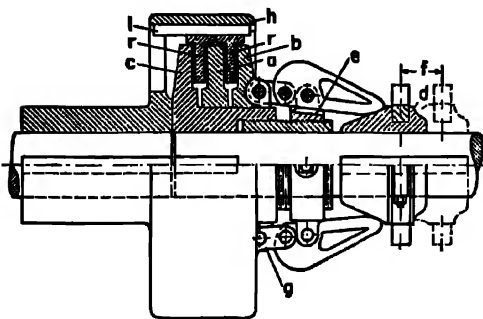


Figure 13. Multidisk clutch. Because of four pairs of contact surfaces it is compact and can be used for comparatively high speeds. The asbestos-lined rings (*r*) can slide on the feather key (*l*) and rotate with the clutch housing (*h*), which is keyed to the left shaft. The inner disk (*a*) and the right disk (*b*) slide on a feather key fastened to the hub of the left disk (*c*), which itself is keyed to the end of the right shaft. When shifting the collar (*d*) to the left in order to engage the clutch, the link (*g*) must be pushed past a position parallel to the shaft axis, thus locking the mechanism. The threaded yoke collar (*e*) serves to take up wear. (Courtesy of International Textbook Company. From Reference 5.)

does not cause failure in the clutch. The principal limitations of these devices are higher initial cost, and unequal wear on mating surfaces due to unequal rubbing velocities.

8.2.3 Special Friction Clutches

In addition to the conventional mechanical clutches used for coupling power trains, a number of special purpose units have system applications:

1. One-way clutches.
2. Centrifugal clutches.
3. Torque-limiting clutches.

One-way clutches, sometimes called overrunning or freewheeling clutches, provide clutching action in one direction only. When rotation is reversed, the device unlocks and ceases to transmit power. The three principal types of the one-way clutch are roller, sprag, and spring types. The roller type contains spring-loaded rollers positioned between adjacent races. They ride on the inner race in a cam-shaped recess (Figure 14). As the races move with respect to each other, the rollers become wedged between them in the narrow part of the cam. When the direction of rotation is reversed, the spring pushes the roller back into the wide part of the recess, and there is slippage between the races.

A variation on this design is the sprag-type clutch. Instead of cylindrical rollers operating in cam-shaped races, this clutch uses cam-shaped rollers

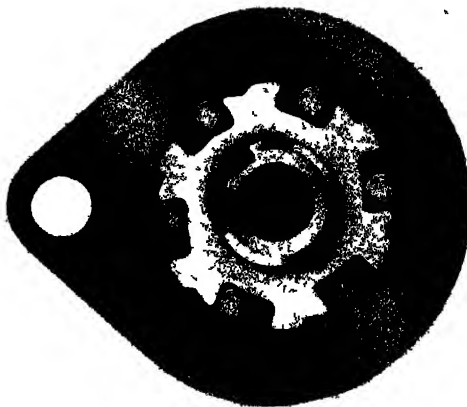


Figure 14. One-way, cam-type clutch (Courtesy of Zero-Max Corporation.)

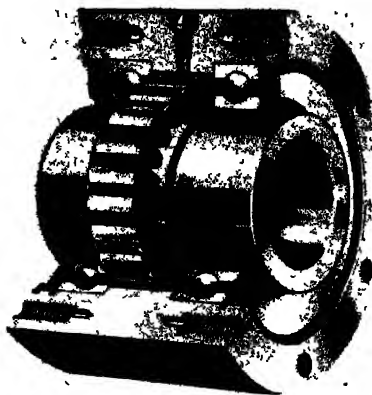


Figure 15. Sprag-type clutch. (Courtesy of Formsprag Corporation.)

operating in smooth races. When the driving element turns, the cam-shaped cylinders rotate and wedge the races together. When the rotation is reversed, the cams turn the other way to unlock the races and permit free-wheeling operation (Figure 15).

Spring-type clutches contain a coil spring wrapped lightly around mating hubs of the driving and driven members of the instrument (Figure 16). When an external torque is applied to the driving hub, the spring tightens on the hub by reducing its inside diameter. It then transmits the driving torque to the other hub as a solid member. Reversing the rotation opens the spring and the clutch slips.

One-way clutches are commonly used in mechanical systems to produce an indexing motion or a checking motion. They can be considered to be the mechanical counterpart of a rectifier or diode. The spring type is used when backlash can be tolerated and when the cushioning properties of the spring are desirable. The roller- and sprag-type units provide quick response. When loads are high, the sprag design is preferred.

A centrifugal clutch allows a driving member to reach a predetermined speed before it starts to drive a load. The most common design consists of an inner driving hub and spring-restrained shoe assembly that rotate inside the driven hub. As the inner shaft speed increases, centrifugal force causes the shoe assembly to expand outward until it reaches the outer hub. At this point the two shafts are clamped together by the shoes, and the clutch allows the driving device to drive the load. This technique is used in electric motors having low starting torque as well as in internal combustion engines (Figure 17).

Torque-limiting clutches are utilized to limit the torque delivered to a

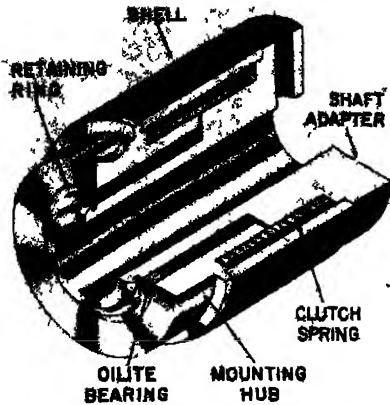


Figure 16. Spring-type clutch. (Courtesy of Marquette Corporation.)

piece of equipment. It is basically a protection against abuse. For example, spark plug torque wrenches can serve to prevent the amateur automobile mechanic from stripping the threads of the spark plug. The device is simply a spring-loaded ball held in a detent that slips out of the recess when the applied torque exceeds a predetermined value. When the load is reduced, the ball is reseated and torque is again transmitted.

8.2.4. Electric Clutches

Most clutches used in systems today are electric clutches, since they provide the control refinements and flexibility that cannot be found in mechanical instruments. The systems application of clutches is very much

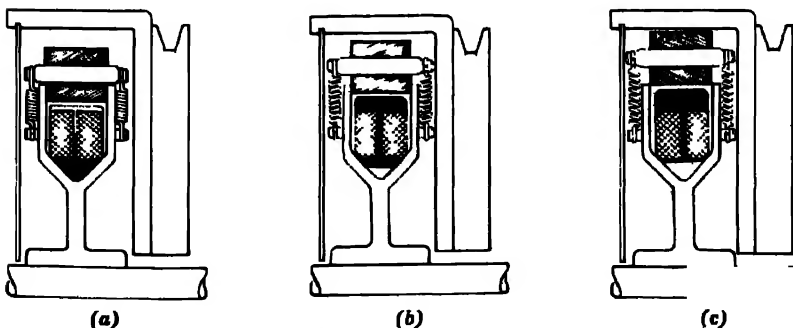


Figure 17. Centrifugal clutch operation. (a) 0-5 sec after motor start, speed 850 rpm; load not engaged. (b) 5-10 sec after motor start; speed 1750 rpm; load not engaged (c) 10-15 sec after motor start; speed 1750 rpm; load smoothly engaged (Courtesy of Mercury Clutch Company)

like the circuit use of a triode or transistor; a relatively insignificant amount of power controls the action of large amounts of power. The current that energizes an electric clutch can be compared to the grid voltage in a triode or the base current in a transistor.

The torque and slip characteristics of a clutch are its most important features. The most efficient use of clutches is at zero slip because the heat generated is minimal and torque is maximal. Three different values for torque can be found in technical literature: static torque, pickup torque, and average torque. Static torque is achieved when the clutch is fully engaged and slip is absent. It is the largest torque that the clutch can transmit. Pickup torque is a value obtained at a given value of slip, normally a value other than that at zero slip. Average torque is the value developed in accelerating a load in a given time interval. Most manufacturers provide torque-slip curves for various heat-dissipation ratings. Very often torque is referred to as slip torque or pickup torque, and slip is called differential speed (Figure 18). Any usage of the device bounded by the axis and the heavy black curve is considered safe. The area to the right of the curve exceeds the thermal rating of the clutch. The heat developed is largely a function of the amount of slip. If slip is an inherent system function, the clutch must be

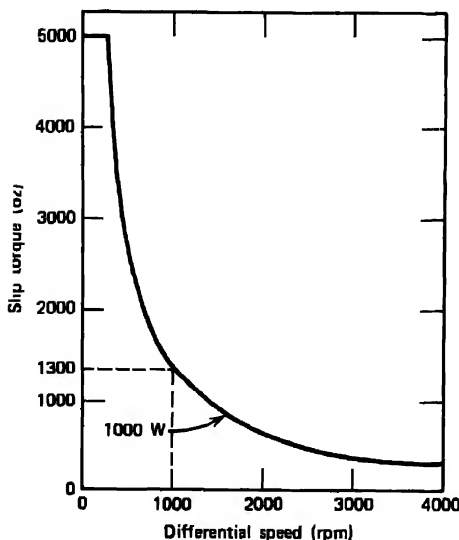


Figure 18. Electric clutch rating curve—average allowable slip torque as a function of slip speed. Dotted lines indicate that for a difference in speed between the input and output of 1000 rpm, the continuous torque must not exceed 1300 oz-in. in order not to exceed the heat-dissipating capacity of this model. (Courtesy of Benwill Publishing Company. From Reference 6.)

selected so that it can provide this type of service safely. An automobile with a standard clutch often fails prematurely when the driver habitually "rides the clutch." By contrast, an automatic transmission equipped with a fluid clutch avoids this type of abuse. As with other devices, heat sinking is extremely important. The rate at which heat is conducted away from the clutch has an important effect on clutch life. Most heavy-duty clutches use some form of air or liquid cooling.

Clutch response time is another important system consideration. In general, the smaller the clutch the faster the response time; therefore it is always a good idea to use the smallest clutch that can handle the load safely. When response times are very critical, some systems utilize "pulsing networks" similar to those described in Section 8.1.4.

There are five principal types of electric clutches available today:

1. Disk type—single and multiple.
2. Tooth type.
3. Hysteresis.
4. Magnetic particle.
5. Eddy current coupling.

The disk-type clutch is the most common electric clutch. Its construction is similar to that of the mechanical clutches already discussed, but it has an actuating solenoid as well. It is relatively cheap, efficient, and available (Figure 19). When the coil is energized, the spring-restrained armature is pulled against the friction plate and remains in the actuated position so long as the coil is excited. Response time is fast, typically 0.005 to 0.015 sec. This design is an "on-off" device and does not work under slip conditions. Since "on-off" service produces little heat, the size of this clutch is quite small for a given torque rating.

When an exceptionally large torque capacity is required in a given envelope, multiple-disk clutches are the choice. They have the same application as single-disk clutches—"on-off" service. Since the power capacity is very high per unit volume, the heat generated is a major design problem. As a result the number of "on-off" cycles per minute is smaller than for a single-disk unit. Most multiple-disk clutches in heavy "on-off" applications have some type of integral cooling system that uses oil or forced air. The most effective of this type of clutch is that which utilizes its high torque capacity with a fairly low-duty cycle.

The electric tooth clutch consists of two toothed plates mounted on concentric shafts that are brought together by solenoid action. It is a positive engagement device and, as a result, produces considerable shock when first engaged. It is not recommended for systems requiring any degree of slip, since the teeth wear badly under partial engagement.

The hysteresis clutch is used when slip is a necessary part of the system

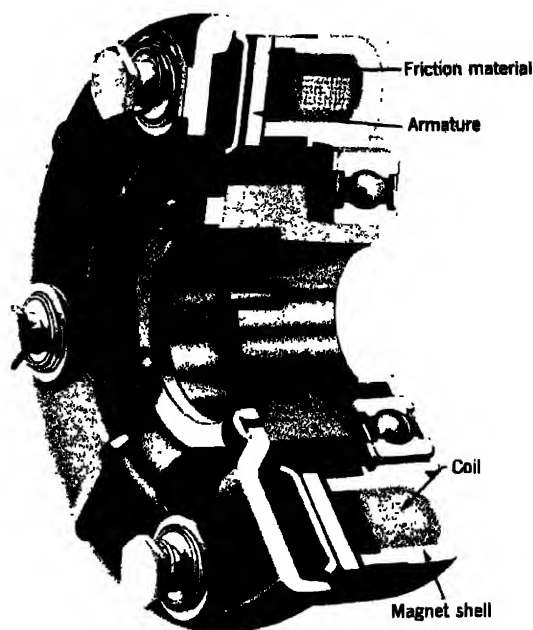


Figure 19. Electric clutch/brake. (Courtesy of Warner Electric Company.)

application (Figure 20). It consists of a stator and coil enclosing two concentric shafts. The input shaft is pressed into a hub equipped with a series of poles to concentrate the magnetic field at discrete points. The output shaft also contains a mating hub that is called a drag cup. The drag cup is of a hard magnetic material with high hysteresis characteristics. When the stator coil is energized, the associated field passes through the input hub and drag cup, forming discrete magnetic poles in each unit, which attract each other. When the input shaft is mechanically caused to rotate, it cuts the exciting field, which in turn causes the electrical position of its magnetic poles to shift forward. The corresponding poles in the drag cup cannot shift so rapidly, since the material has a high hysteresis characteristic. As a result, the drag cup (output shaft) always lags behind the input shaft. This is the desired slip characteristic. The degree of slip is a function of the control excitation of the winding, and the hysteresis lags in the drag cup. Since the hysteresis is essentially constant, the slip is easily controlled by varying the coil current. A typical torque-slip curve is shown in Figure 18. Some designs use permanent magnets to supply the magnetic flux. This

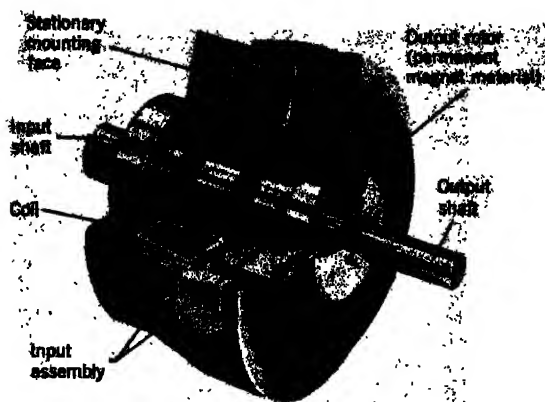


Figure 20. Hysteresis clutch. (Courtesy of Magtrol, Inc.)

results in smaller, less expensive units, but the torque level cannot be adjusted. Clutches with a conventional winding can be adjusted to function from near-synchronous speed to a wide range of slip speeds. Torque is independent of speed. The heat generated is a critical item, and it must be removed or the unit derated. This clutch is used extensively in small precision instruments and aerospace projects where cushioning shock loads, low power consumption, and long life are important. Hysteresis clutches are less attractive when cost and high power capacity are prime considerations.

The closest approach to a universal instrument clutch is the magnetic particle clutch. It combines wide torque range, fast response time, torque-independent-of-speed characteristics, and smooth starting torque with easy availability. As in the case of hysteresis clutches, the price is higher than for friction units. Compared to hysteresis units it has a greater torque range and faster response. Torque is independent of slip speed.

The input and output shafts are equipped with concentric hubs separated by an air gap (Figure 21). The gap is filled with an extremely fine, dry magnetic powder. A coil is built into the housing to create a magnetic field in the gap. This causes the powder to form a rigid link between the input and output shafts. The torque transmitted is a function of the magnetic field strength in the gap; consequently, the slip can be controlled by adjusting the coil excitation. When the clutch slips, considerable heat is generated by the internal rubbing action of the magnetic particles. The amount is a function of speed, torque, and slip. The most common approach to this problem is to provide a water jacket in the stator. Another variant of the magnetic clutch is the use of a fluid loaded with magnetic particles forming the

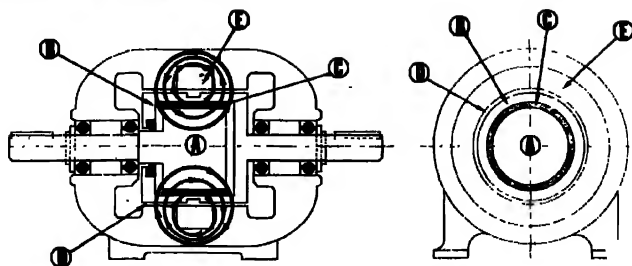
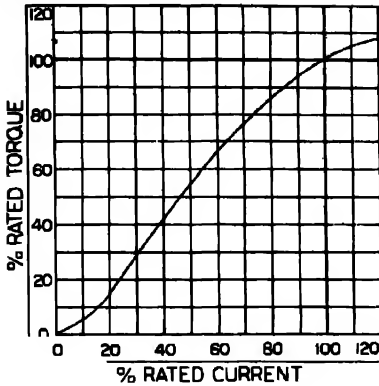


Figure 21. Magnetic particle clutch. The two moving parts—the inner rotating member (A) and the outer rotating member (B)—are cantilevered on bearings and do not touch each other. The space between the inner and outer rotating members is called the working gap (C). This gap is filled with an extremely fine, dry magnetic powder. There is a second gap (D) between the outer rotating member and the stator frame. A stationary 90 V DC coil (E) is built into the stator frame around this area. When current is passed through the coil, a magnetic field is created, causing the powder to form a link between the inner and outer rotating members, thus transmitting the torque. The amount of torque transmitted varies as the magnetic field strength is increased or decreased through the change in current. (Courtesy of Vickers Division, Sperry Rand Corporation.)

magnetic coupling. This has the advantage of smaller internal friction and smoother operation. The disadvantage is that seals must be provided to prevent leakage with consequent cost and wear complications. The response time is excellent; smaller units are capable of providing 3 to 5 msec response. This is due to the low inertia of the output shaft and the lack of mechanical motion required to engage or disengage the device. The control of torque transmitted is particularly convenient in this device (Figure 22). The torque is completely independent of slip down to zero speed. Typical operating temperature limits are 125°C with a life of 5×10^6 cycles at 60 Hz. Typical applications include time-dependent servomechanisms and high-inertia loads such as printing presses, elevators, and machine tools.

The eddy current coupling is sometimes used as a clutch, but more often it serves as a variable speed drive with electric motors. Unlike the hysteresis or magnetic particle clutches, torque increases with slip. The device consists of an input drum, field coil, and coupling pole assembly or output shaft. The field coil is built into the stator of the clutch and encompasses the input and output drums. The pole assembly is designed to direct the flux so that certain areas form high-flux densities when the coil is energized. When the input drum rotates, eddy currents are induced on its inner surface. These eddy currents create a new magnetic field, which interacts with the field in the pole assembly, causing the pole assembly to rotate. The drum must always rotate faster than the pole assembly to create coupling torque. The torque transmitted is a function of the square of the

MAGNETIC PARTICLE CLUTCH



MAGNETIC PARTICLE BRAKE

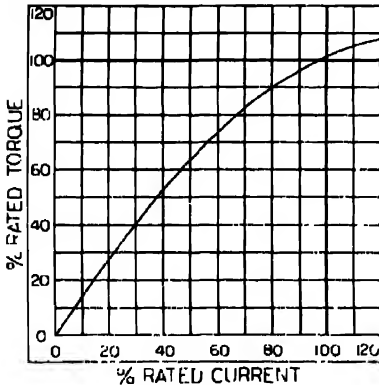
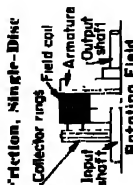


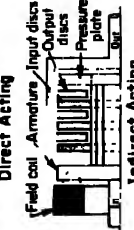


Figure 22. Magnetic particle control curves. Note that these are composite curves. (Courtesy of Vibrac Corporation.)

field strength and the slip between input and output shafts; therefore there can be no output at zero slip. This may occur at standstill or synchronous speed. The eddy current losses in this device produce a substantial amount of heat, and the device usually is designed with some form of cooling jacket. Low-duty units may use air passed over the hot surfaces, but heavy-duty units use water jackets. Permanent-magnet eddy current couplings are available, but they do not have the torque capacity or flexibility of conventional units. This instrument is useful where very large torque must be accommodated and response time is slow. A summary of clutch characteristics is given in Figure 23.

Type	Rating	Advantages	Limitations	Typical Applications
Friction, Single-Disk  Collector rings, Field coil, Input shaft, Armature, Output shaft, Rotating Field	Torque: 1.5 to 1350 lb.-ft. Diameter: $\frac{1}{4}$ to 15.1 in. Speed Range: 0 to 10,000 rpm. Response Time: 9.5 to 280 millisecc.	Fast-acting; high torque-to-size ratio; high power amplification; low cost; torque easily controlled; available in wide range of sizes; excellent duty-cycle capabilities.	Some wear on friction surface, requiring adjustment or replacement of friction faces; must be derated to operate at continuous slip.	Suitable for either on-off or slip. Used in industrial and commercial equipment, such as: Machine tools, business machines, textile machinery, packaging equipment, material handling, farm equipment, construction equipment, automotive accessories. Miniature precision devices also used for aerospace and instrument applications.
Friction, Multiple-Disk  Field coil, Input discs, Output discs, Armature, Stationary Field	Torque: 28 to 1200 lb.-ft. Diameter: 4 to 10 in. Speed Range: 0 to 4000 rpm. Response Time: 28 to 200 millisecc.	High torque-to-diam ratio; operates in oil atmosphere; long life rugged.	Limited heat-dissipation capability; limited duty-cycle rate; should be used in oil atmosphere.	Direct-acting type is used where accuracy is primary consideration, as it automatically compensates for wear. Indirect-acting type used where life is primary consideration. Applications for both types include speed-changing equipment and transmissions for multiple-spindle chucks, automatic screw machines, and boring machines.
Direct Acting  Field coil, Armature, Input discs, Output discs, Pressure plate				
Indirect Acting  Field coil, Armature, Input discs, Output discs, Pressure plate				

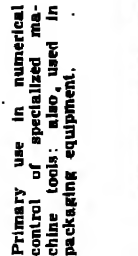
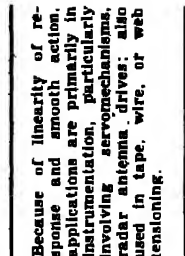
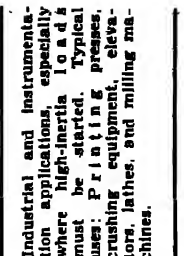
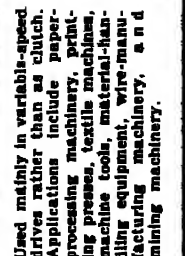
	<p>Torque: 75 to 1200 lb.-ft. Diameter: $3\frac{1}{4}$ to 6$\frac{1}{2}$ in. Speed Range: 0 to 4500 rpm.</p>	<p>Positive engagement high torque transmission; no idling torque; operates in oil atmosphere.</p>	<p>Can be engaged only at low slip; engagement shock.</p>	<p>Primary use in numerical control of specialized machine tools; also, used in packaging equipment.</p>
	<p>Torque: 6-1800 oz.-in. Diameter: $\frac{3}{8}$ to $6\frac{1}{4}$ in. Speed Range: 0 to 15,000 rpm. Response Time: 48 millisecc to 2.2 sec.</p>	<p>Torque independent of speed; high torque-to-signal linearly below saturation (except near zero); smooth operation; long life and little wear if not overheated; environmentally stable.</p>	<p>High cost; low power gain; limited to low-torque applications; torque derating required at high slip speeds.</p>	<p>Because of linearity of response and smooth action, applications are primarily in instrumentation, particularly involving servomechanisms, radar antenna drives; also used in tape, wire, or web tensioning.</p>
	<p>Torque: 18 oz.-in. to 600 lb.-ft. Diameter: $1\frac{1}{2}$ to 10$\frac{1}{4}$ in. Speed Range: 0 to 12,000 rpm. Response Time: 5 to 16 millisecc.</p>	<p>Torque independent of speed; high torque-to-signal linearly below saturation (except near zero); smooth operation; quick response; high torque-to-inertia ratio; low excitation.</p>	<p>High cost; limited heat-dissipation capability; limited life; torque derating required at high slip speeds.</p>	<p>Industrial and instrumentation applications, especially where high-inertia loads must be started. Typical uses: Printing presses, crushing equipment, elevators, lathes, and milling machines.</p>
	<p>Torque: $\frac{1}{2}$ to 10,000 lb.-ft. Diameter: Not applicable; minimum housing size $5\frac{1}{2}$ in. Speed Range: 33 to 3000 rpm. Response Time: 0.38 to 1.16 sec.</p>	<p>Torque proportional to slip speed; square-law torque-to-signal characteristics; smooth operation; high power gain; indefinite life.</p>	<p>High cost; no torque at zero slip; moderately slow response; high losses during steady operation with friction load; must be used at low slip for efficient operation; temperature-sensitive—may require water-cooling.</p>	<p>Used mainly in variable-speed drives rather than as clutch. Applications include paper-processing machinery, printing presses, textile machines, machine tools, material-handling equipment, wire-manufacturing machinery, and mining machinery.</p>

Figure 23. Characteristics of electric clutches. (Courtesy of Machine Design Magazine, Penton Publishing Co., Cleveland, Ohio, Dec. 18, 1969. From Reference 4.)

In addition to the electromechanical clutches discussed, air clutches are widespread. The same mechanisms can be utilized. The basic difference is that air pressure, instead of mechanical thrust, forces the clutch elements together. The advantages of air clutches are very high torque ratings, fast response, and smooth operation. The disadvantage is that auxiliary piping and a source of regulated compressed air is required. Typical performance curves are shown in Figure 24.

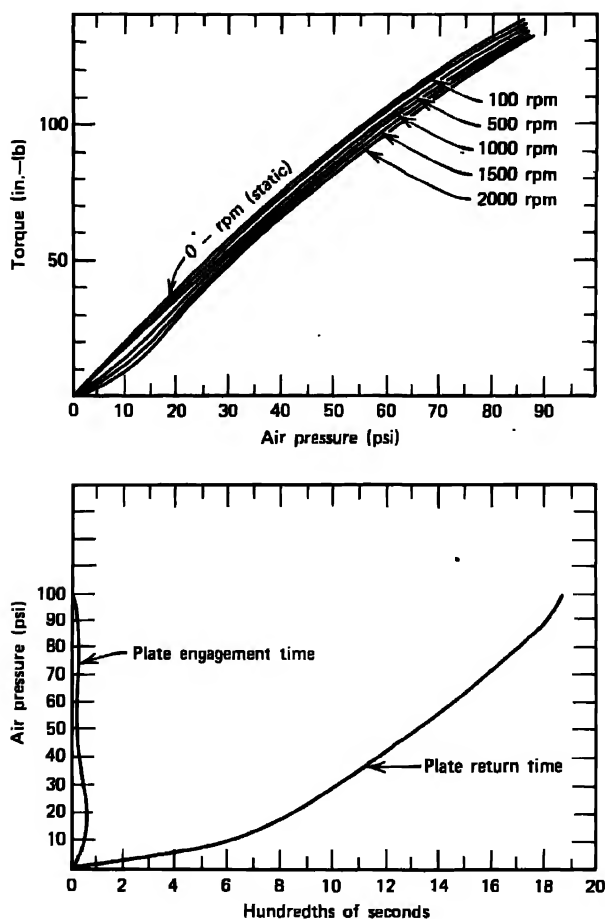


Figure 24. Air clutch characteristics. (Courtesy of the Horton Corporation.)

8.3. BRAKES

As previously noted, brakes can be thought of as a special form of clutch that has one of its shafts rigidly fixed. Each generic type of clutch has an equivalent brake design. The first brakes adapted for systems applications used simple friction-type brakes modified so that a solenoid replaced a lever or crank to actuate the device. Examples of early mechanical designs are shown in Figure 25*a* and *b*. The automotive-type brake is extremely popular since it is easy to obtain and readily adaptable to solenoid control (Figure 25*c*). Modern friction brakes have integral magnetic devices to force the friction surfaces together. They are commonly called magnetic brakes. A single friction disk brake is shown in Figure 26. When the armature is energized, the armature assembly is pulled against the friction disk; when power is removed, the spring forces the armature back to its initial position. When greater capacity is required, multiple disk brakes are useful (Figure 27). This design has a fail-safe provision—power is required to release the brake. The relative merits of the various brakes are similar to those listed for the equivalent clutches.

The most commonly used brakes in modern system design are the hysteresis and magnetic particle brakes, which are identical to the clutches shown in Figures 20 and 21.

Clutches and brakes are normally used together in many applications. Some of the common configurations and applications are shown in Figures 28*a* and 28*b*. A clutch and brake combination in one package requires less space than do separate components. System reliability is improved, since the number of connections is minimized. The backlash between components is also reduced. The cost of a package is normally lower than that of separate components.

8.3.1. Brake and Clutch Selection

The same criteria are used in selecting clutches and brakes. Since any system using clutches and brakes consists of a series of cycles, the first step is to define the cycle. A typical example is shown in Figure 29. It is important to specify realistic response times that can be realized with shelf hardware. It is also important to realize that the cycle is an idealized curve and is subject to modulation by the following occurrences:

1. Backlash and slippage between mating parts.
2. Variations in stiction and friction as a function of wear and contamination.
3. Eccentricity and misalignment of centerlines of rotating parts.

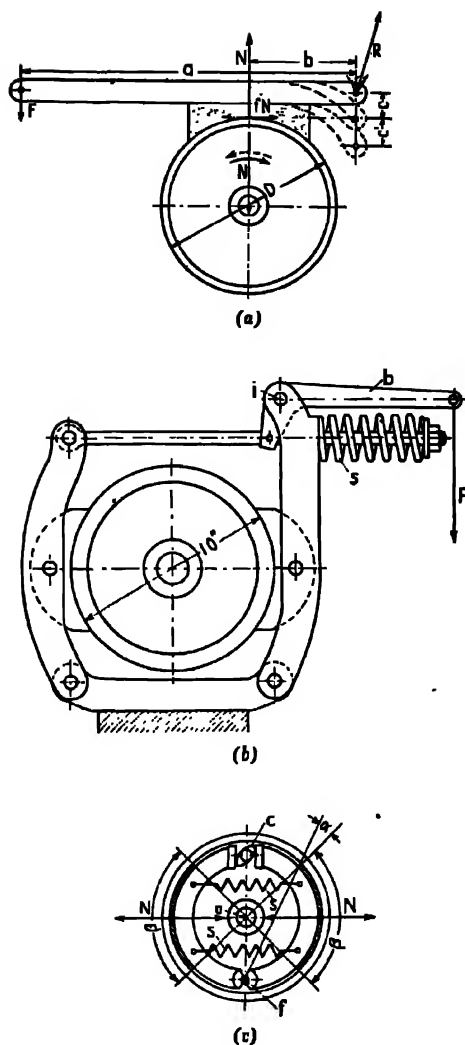


Figure 25. Mechanical brakes. (a) Single-lever block brakes. (b) Double-block crane brake. (c) Internal block brake. (Courtesy of International Textbook Company. From Reference 5.)

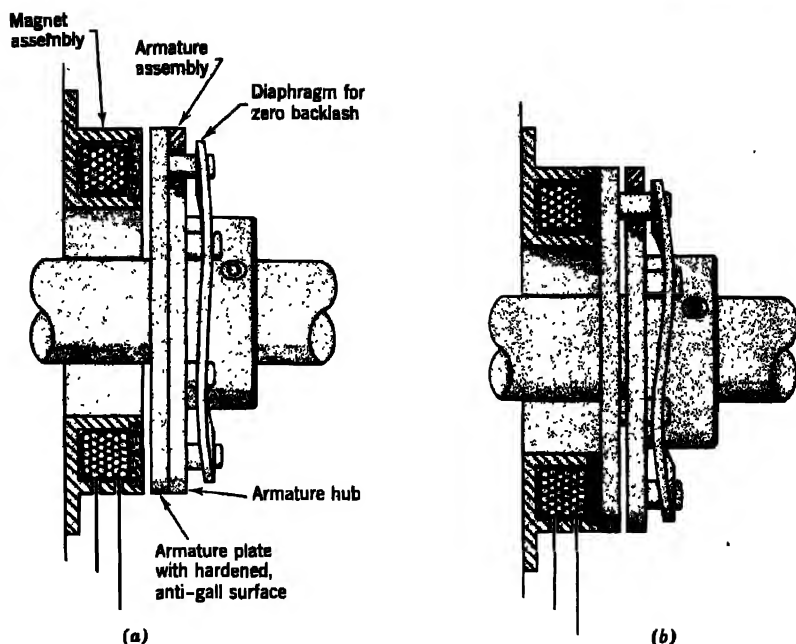


Figure 26. Magnet brake. (a) Deenergized. (b) Energized. (Courtesy of Simplatrol Corporation.)

The second step in selection is to calculate the torque rating. There are two principal formulas:

$$T = \frac{HP \times 5250}{N} \times SF \quad (2)$$

where T = torque (foot-pounds)

HP = horsepower transmitted

N = speed (revolutions per minute)

SF = safety factor, normally 1.5 to 5.0, depending on the duty cycle and anticipated overload conditions

and

$$T^1 = \frac{WR^2 \times \Delta N}{308 \times t} \quad (3)$$

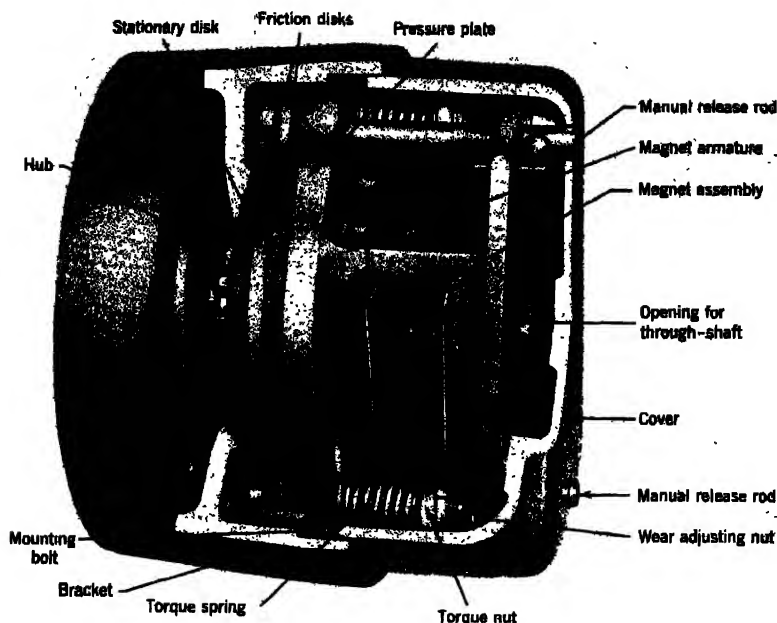


Figure 27. Magnetic multidisk brake. (Courtesy of the Dings Corporation.)

where T^1 = torque required to accelerate or decelerate the load; the torque required to overcome friction must also be added to this figure

W = weight of the load and clutch-brake unit (pounds)

R = radius of gyration of the unit (feet)

ΔN = change in speed during the acceleration or deceleration (revolutions per minute)

t = time required for starting or stopping (seconds)

The first formula should be used only if it is impossible to estimate the inertial load of the system. The second expression results in a more accurate estimate of the actual torque required. Pulsing networks can be used successfully to reduce response time.

The next step is to check the static and dynamic torque curves of the instruments being considered (Figure 30). The static torque curve shows the relationship between the output torque and applied voltage when there

is no relative motion between adjacent faces of the clutch or brake. The point at which the unit will start to slip for a given excitation is indicated. The dynamic torque curve shows the torque available to pick up or stop at a given speed differential between mating surfaces. If the torque of a clutch is not sufficient to overcome the friction load it will never accelerate, but will slip indefinitely. In the case of a brake, friction assists stopping. The curve shown is for 100% excitation; for higher or lower voltages a series of parametric curves would appear on the chart.

It is useful to calculate the amount of heat generated for a given cycle so that heat transfer computations can be made:

$$E = 0.00017(WR^2)N^2F \quad (4)$$

where E = heat generated (foot-pounds per minute)

W = weight of the load and clutch-brake unit (pounds)

R = radius of gyration of the load and clutch-brake unit (feet)

N = speed (revolutions per minute)

F = number of starts and stops per minute

If a choice is possible, it is best to locate the clutch on the low-speed side of a gear train. This location requires a larger clutch than does the high-speed side, but offers the following advantages:

1. The clutch has longer life because of greater heat-dissipating ability.
2. The components between the motor and the clutch, such as the speed reducer, are not so subject to shock as are the clutch and the brake.






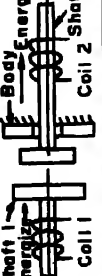

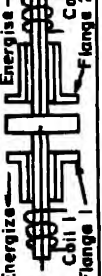


The excitation voltage used on electric brakes and clutches should be DC whenever possible for the same reasons as given for solenoids. (Section 8.1.3). Although filtered AC voltage is sometimes used, the filtering done to convert AC to DC power leads to delays. It may also cause chatter in the unit.

8.4. BEARINGS

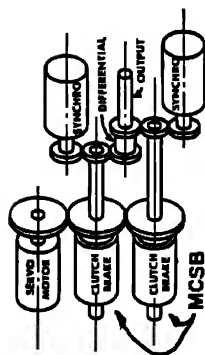
This section discusses some of the new bearing trends rather than standard machine design applications. Prime attention is given to practices that supplement standard bearing design.

Today two industries stimulate progress in technology: the space industry and automation. Both fields require a degree of reliability that was considered almost impossible a few years ago; the big difference between the two areas is cost. In space work, cost is no deterrent, whereas the automation industry is very price-conscious.

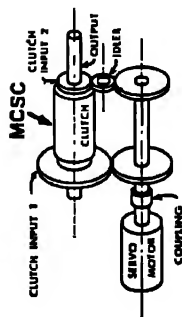
TYP CON		FRI FRI		REL LED	DESCI		APPLCATH	20
	Static Output Torque		MC	Magnetic Clutch De-energized: Energized:	- Input and output shafts free Input shaft coupled to output shaft	E		
	Static Output Torque		SC	Spring Clutch De-energized: Energized:	Shafts clutched (by spring) Shafts free	J		
	Static Output Torque		MB	Magnetic Brake De-energized: Energized:	Shaft free Shaft braked to body	F, L		
	Static Output Torque		SB	Spring Brake De-energized: Energized:	Braked Free	I		
	Static Output Torque		MSCB	Magnetic Clutch — Spring Brake De-energized: Energized:	Input free, output braked to body Input clutched to output	A, C		

		SCMB	Spring Clutch — Magnetic Brake De-energized: Input clutched to output Energized: Output braked to body, input free	C
		MSC	Magnetic Clutch — Spring Clutch De-energized: Flange 1 clutched to shaft, flange 2 free Energized: Flange 2 clutched to shaft, flange 1 free	B, H
		MCMB	Magnetic Clutch — Magnetic Brake De-energized: Both shafts run free Coil 1 Energized: Shaft 1 clutched to shaft 2 Coil 2 Energized: Shaft 1 free, shaft 2 braked Both Coils Energized: Both shafts braked	K
		MCMB (duplex)	Magnetic Clutch — Magnetic Clutch De-energized: Shaft and both flanges run free Coil 1 Energized: Flange 1 clutched to shaft, flange 2 free Coil 2 Energized: Flange 1 free, flange 2 clutched to shaft	D, G
		MCMB (triplex)	Magnetic Clutch — Magnetic Clutch De-energized: Shaft and both flanges run free Coil 1 Energized: Flange 1 clutched to shaft at "A", flange 2 free Coil 2 Energized: Flange 1 free, flange 2 clutched to shaft at "B" Both Coils Energized: Both flanges clutched to shaft	No application shown

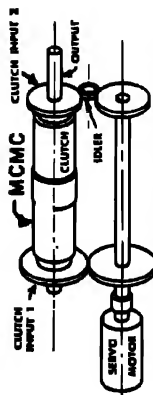
(a)



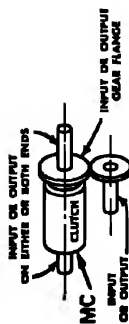
A — Clutches used to add or subtract two inputs.



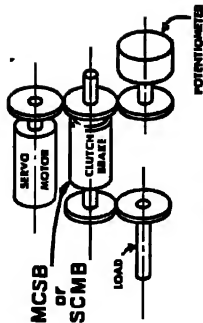
B — Direction changing.



D — Direction changing and uncoupling.

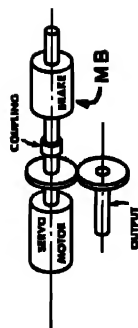


E — Coupling or decoupling power or sensing.

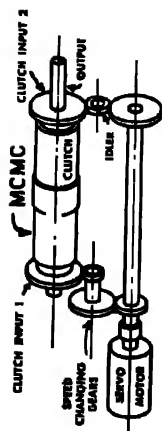


C — Used to adjust a load.

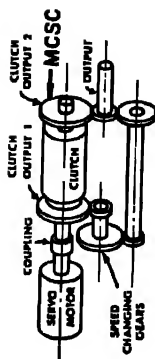
1. Clutch engages, motor drive load to position left or right.
2. Brake engages, holds load.
3. Servo and pot return to neutral.



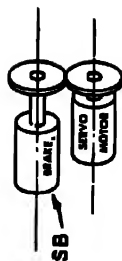
F — As a motor brake.



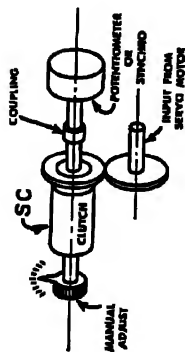
G — Speed changing and uncoupling.



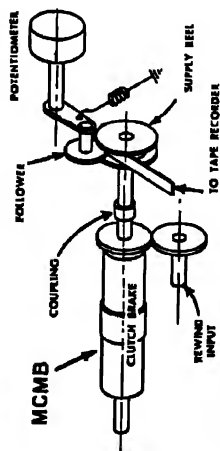
H — Speed changing.



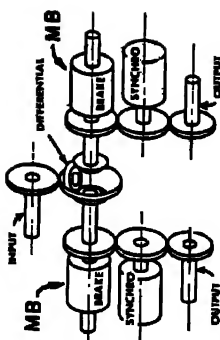
I — Normally braked, energise to free, wire clutch in series or parallel with motor.



J — Energise to adjust.



K — Used as a constant tension device.



L — Controlling output from differential.

(b)

Figure 28. (a) Clutching and braking devices and their coupling characteristics. (b) Application schematics of clutch-brake devices. Note that units are identified by letters defined in (a). (Courtesy of Benwill Publishing Company.)

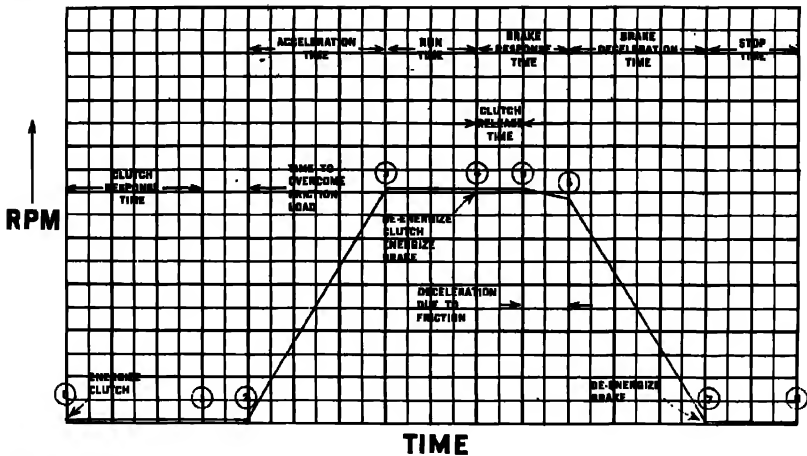


Figure 29. Duty cycle for brake-clutch unit. (Courtesy of Simplotrol Corporation.)

To describe the new developments in each field, we briefly review their backgrounds.

8.4.1. Space Environment

Space environment has five primary hazards:

1. Large temperature variations.
2. Low pressure.
3. Radiation fields.
4. Space debris.
5. Zero gravity.

Ambient temperature in space may be close to absolute zero. However, because of heat generated by satellites, bearings and other components seldom become colder than -250°F . The two modes of heat transfer in space are conduction and radiation. Any heat generated by a bearing must be transferred to a heat sink to prevent overheating. Similarly, any heat generated by motors, torquers, or other instrumentation must be channeled to the heat sink so that it is not transmitted to temperature-sensitive components. Even a few watts of power transmitted continuously to a given element will cause serious thermal problems if dissipation does not occur. The prime source of heat in space is, of course, the sun. Radiant energy has been estimated to be $442 \text{ Btu/hour/ft}^2$; it is therefore possible

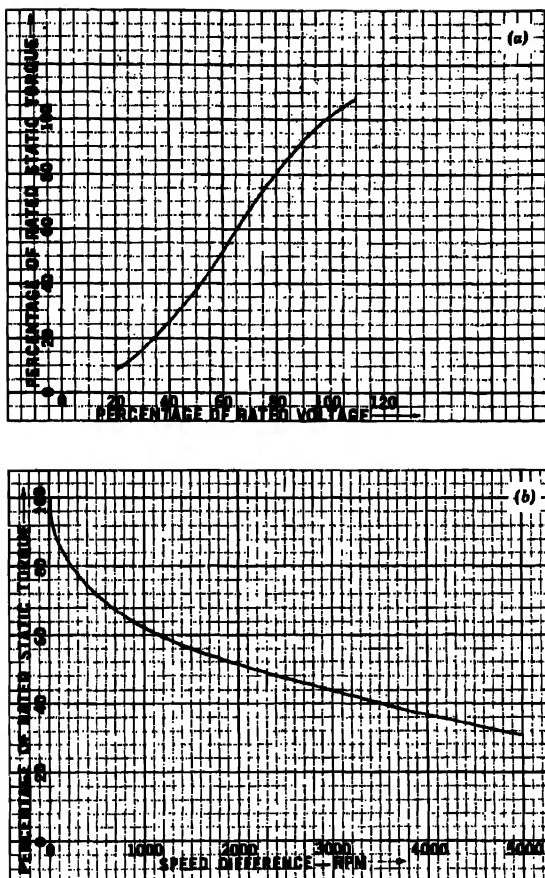


Figure 30. (a) Static torque curve for a brake-clutch unit. (b) Dynamic torque curve for a brake-clutch unit. (Courtesy of Simplatrol Corporation.)

for components in orbit to be stabilized at -250°F and then, upon exposure to sunlight, to be rapidly heated to 250°F or above. Because of the temperature shock involved, components must be designed not to fail during the transition.

Lubrication of bearings in this environment becomes very difficult. At the low pressures, the lubricant may vaporize. When the lubricant disappears, the load-carrying capacity of a bearing is greatly diminished. Relatively small changes in temperature will cause changes in lubricant

viscosity that represent large variations in viscous friction, thereby complicating servo control problems.

The low pressure in space greatly accelerates the volatilization of lubricants in the bearings. Another problem is the variation in level of friction with atmospheric pressure. Since one of the functions of lubricants is to carry away heat, local hot spots may develop on the bearings.

Of all the radiation in space, ultraviolet rays present the most serious problem for bearings. The danger is their ability to change the surface structure of some materials. Epoxies are particularly affected. In other materials, the emissivity is changed. This may result in overheating of components.

The principal hazard of meteoroids is from particles that range in size from 10^{-3} to 10^{-13} gram. A significant quantity of them are attracted by the earth's gravitational field and may cause erosion-type damage. Surface finishes may deteriorate, which in turn affects the ability of a surface to radiate heat. Another effect is actual penetration of surfaces. Spalling, or flaking, of material may also occur.

The effect of zero gravity is significant where pressure lubrication is not available, since the conventional flow of lubricant to a sump does not take place.

Many space-type bearing problems have been solved by providing the vehicle or subassembly with a molecular flow seal (labyrinth seal). This technique consists of establishing a long path that molecular components must follow before reaching a low pressure sump. A substantial pressure can thus be maintained in the bearing. If this pressure is greater than the vapor pressure of the lubricant, there will be no vaporization. An improvement on this technique is to provide a small tank of gas that bleeds into the bearing area through an orifice. The pressure can thus be maintained for a much longer period of time.

8.4.2. Dry-Film Lubrication

Keeping a lubricant in the bearing is a major problem. In the past few years several greases and oils have been developed with low enough vapor pressures to withstand the space environment for a reasonable period of time. Silicone greases are examples. A second technique is to use a dry-film lubricant. This method exhibits the following desirable characteristics:

1. Resistance to temperature extremes.
2. Compatibility with many fuels and oxidizers.
3. Resistance to hard vacuums.
4. Maintenance-free service life.
5. High static load capacity.

Dry-film coatings provide a low-friction interface between two mating metals to prevent wear. This is usually accomplished by impregnating the bearing retainers with a suitable lubricant such as molybdenum disulfide, tungsten disulfide, or PTFE (polychlorotrifluoroethylene). A thin coating, typically about a mil, is sufficient for most applications. Some designs provide recesses in the retainers where reservoirs of PTFE may be stored; the rubbing action of the balls on the lubricant soon coats all rotating surfaces of the bearing with the substance.

Dry-film coatings were originally developed for slow-speed, high-load, sliding motions; journal bearings, spherical bearings, gimbal mountings, and ball joints for jet engine mountings are typical examples. Rolling contact bearings were not considered good applications because of short wear life and high friction due to wear debris. Recent developments in bearing technology have corrected the situation and are discussed later in this book.

Some significant design guides for utilization of dry-film lubricants are the following:

1. Do not rely on dry-film coatings to provide corrosion protection.
2. Do not use a supplementary lubricant in conjunction with dry-film lubricants, since the latter will contaminate the coated surface and cause loss of the primary lubricant.
3. The design should prevent the coated surface from moving in and out of the load zone. This coated surface becomes burnished during use and usually is flaky. It is more easily removed, in this condition, when moved in and out of the load zone.
4. Remove all sharp edges that can damage the dry film during assembly.
5. Coating both rubbing surfaces leads to the best performance.

One criticism of dry-film lubrication has been that, because of debris from lubricants, it produces higher frictional torques than competitive methods. However, these techniques are being refined very rapidly and should be investigated for each problem.

Dry-film lubrication prevents the worst type of failure a bearing can experience—cold welding, or, more accurately, adhesion or cohesion of adjacent parts.

Several studies have tried to determine the effect of the space environment on metal adhesion and cohesion. Adhesion is defined as the sticking together of surfaces of different, separate materials; cohesion is the sticking together of surfaces of the same material. Both processes may occur when clean mating surfaces are brought sufficiently close together to permit interatomic forces to be effective. The important factors are pressure, time, and temperature. Pressure causes the mating surfaces to come into intimate contact. It

also causes high spots on either surface to absorb the initial load and deform before the entire area receives the load. Increased temperature decreases the strength of materials, hence the force required for intimate metal-to-metal contact.

Oxides or surface films prevent metal-to-metal contact and must be removed before bonding can occur. They can be destroyed in three ways:

1. Diffusion of the oxide into the base material.
2. Mechanical rupture of the brittle oxide.
3. Dissociation of the oxide.

The time needed to reform a surface film after it has been removed is much longer in a vacuum than in air. However, the requirement for a clean surface is more likely to be found in a vacuum than in air.

The Miniature Precision Bearing Company has developed a dry lubricant called Dicronite that has a great variety of applications including roller and ball bearings. The process applies a fusion-bonded film of modified tungsten disulfide to a metal substrate in the retainer. The lubricant is not free to peel off in the bearing, as were earlier dry lubricants. The film thickness is about 20 μ in. The modified tungsten disulfide is inert, insoluble, and stable up to 900°F in air and 2400°F in a vacuum. It has also been used successfully in cryogenic applications. The associated coefficient of friction is 0.025 to 0.090. Dicronite is envisioned as a technique for lubricating machinery that formerly had to operate without the benefit of conventional lubricants because of the rate at which these normal materials oxidize or contaminate the atmosphere. It is also used in combination with conventional lubricants to provide dry-film lubrication when conventional lubrication is ineffective. An example is high-acceleration turbine machinery where the oil pump is driven from the main shaft.

The Barden Corporation has also perfected a dry bearing using Bar Temp that is designed to function over a temperature range of -325 to 575°F. Bearing rings and balls are made of 440C stainless steel, and a precisely machined one-piece cage serves as a ball separator and a source of dry lubricant, depositing a light coating on the raceways. Bar Temp bearings do not contain any lubricant, in the normal sense; consequently, there is no problem with variation of oil viscosities at temperature extremes. The basic design is capable of sustaining end thrust in both directions as well as radial load. Typical applications have been in aerospace motors, servometers, actuators, and gyros. An unusual application was as a lubricant in a pump operating in an atmosphere of hydrogen and water vapor. This environment normally washes out conventional lubricants. Bar Temp survived more than 1000 hours on initial tests without degradation of performance (Figure 31).

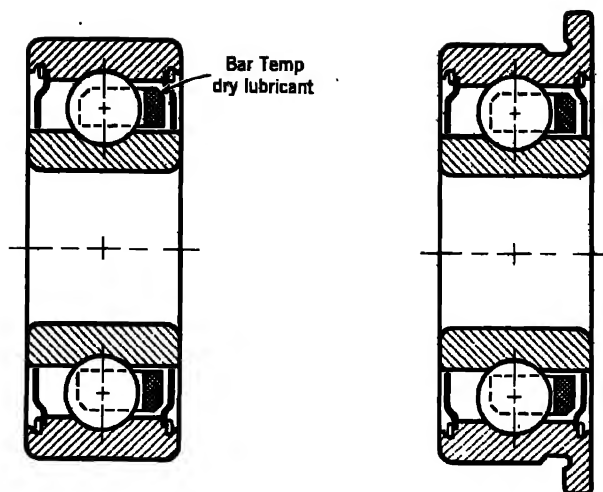


Figure 31. Bearing with dry lubricant. Cross-sectional view of BarTemp deep groove bearings, flanged and unflanged, shielded with a one-piece retainer. (Courtesy of Barden Corporation.)

8.4.3. Barrier Films

In many instruments migration of lubricants from the intended area to other areas of the mechanism has produced severe degradation of performance. In some applications the loss of lubricant leads to bearing failures due to partially dry bearings; other devices fail because the lubricant is itself a contaminant and may also carry other debris with it. In synchros, for example, silicone lubricants will contaminate slip rings and destroy the electrical integrity of the instrument. Barrier film techniques were developed to keep lubricants in the areas where needed. In its simplest form, barrier film is a coating of fluorinated methacrylate applied to all areas in the path of the lubricant. The areas to be coated are, typically, the external surfaces of bearing races, the shield, retaining ring, and the shaft in the area of the bearing. The coating effectively provides a barrier to the flow of lubricant out of its intended area of action. Exact techniques of application, safe solvents, and the best barrier materials are currently being developed. The Navy Ammunition Depot at Crane, Indiana, reported that barrier films considerably extended the storage life of synchros. In its work with barrier film, the MPB Corporation reports, conventional gyrospin bearings using impregnable phenolic linen and phenolic paper base retainers showed 25 to 50% life extensions.

Examination of the motors after a life test showed that barrier-film-treated bearing assemblies successfully contained excesses of lubricant which were present within the bearing envelope and did not permit the oil to migrate and contaminate the rest of the instrument. Also being investigated are bearings without reservoir-type retainers, where the lubricant is held in the areas solely by barrier films. Possibly the application of barrier films to the faces of gyrospin bearings will be sufficient to keep the lubricant in the bearings. It has been established that 100 to 200 μg of oil is sufficient for thousands of hours of operation of a typical gyrospin bearing of the R4 size. This quantity of oil can be maintained on bearing rolling element surfaces by barrier films without the need of reservoir-type retainer material.

8.4.4. Cryogenic Bearings

Cryogenic bearings were developed to meet the need for a bearing capable of functioning in a cryogenic fluid. At low speeds and light loads, stainless steel bearings with phenolic separators worked moderately well. As speeds and loads increased, Teflon or graphite compositions were used in the separators to provide improved self-lubrication characteristics. The next stage of development was to use the cryogenic fluid as the lubricant. The basic differences between conventional and cryogenic lubricants are temperature and viscosity. Cryogenic temperature increases ultimate strength and hardness of metals, but it also increases brittleness. Materials that are not acceptable for rolling elements at room temperature can function at very low temperatures.

The viscosity of liquid hydrogen is about 0.013 cP while the average conventional lubricant is 10 cP (Figure 32). The result is lower viscous drag and less damping than in conventional bearings, as well as smaller capacity for supporting combined radial and thrust loads at high speeds. The failure mode of cryogenic bearings, it has been determined, is due to progressive loss of internal clearances and subsequent destruction of raceways, balls, and separators. The primary cause of dimensional changes is internal heat generated within the bearing.

Heat generation in any high-speed bearing emanates from three primary sources:

1. Spinning and rolling motion at the race contact areas.
2. Separator rubbing friction.
3. Churning of the lubricant in the bearing.

Spinning and rolling of the balls on the raceway occur in any angular contact bearing; spinning is more pronounced at high speed because centrifugal force causes dual-contact angles which increase the rate of

Gas	°F
Acetylene	-119.2
Air	-317.9
Argon	-372.6
Carbon Dioxide	-109.3
Ethylene	-154.8
Fluorine	-306.2
Freon 13	-114.6
Helium	-452.1
Hydrogen	-422.99
Methane	-258.6
Nitrogen	-320.5
Oxygen	-297.3

(a)

Average lubricating oil	=	10-1200 (at 20°C)
Water	=	0.98 (at 20°C)
Liquid nitrogen	=	0.158
Liquid oxygen	=	0.190
Liquid hydrogen	=	0.013

(b)

Figure 32. Properties of gases at cryogenic temperature. (a) The normal boiling point at 1 atm pressure or the point at which some of the more frequently used gases are liquefied. (b) The dynamic viscosity of cryogenic fluids compared to oil and water (centipoise). (Courtesy of ITI Corporation. From Reference 15.)

spinning, friction, torque, and heat generated. The heat generated by this action must be dissipated by the lubricant. The introduction of sufficient amount of cryogenic fluid into the bearing is the best way of lubricating and cooling the elements of the bearing. Separator designs with minimal restrictions to flow and minimal friction are required. Materials used are Teflon and graphites. Since they are structurally weak, they are reinforced with aluminum sideplates. Armalon and Rulon (teflon-filled materials) are also utilized. Typical cryogenic bearings are shown in Figure 33.

Cryogenic materials have application in the steel industry as well as in space. Handling of nitrogen, argon, hydrogen, and oxygen in liquid form provides a great saving. Another nonmilitary application is in chemical processing. In space work the major field for cryogenic machinery is pumping fluids. In these pumps bearings operate completely submerged in the

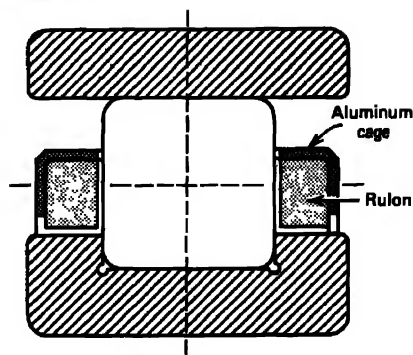


Figure 33. Cryogenic bearing. Cross section showing Rulon separator and aluminum cage. The separator construction is such that all sliding surfaces both in roller pockets and on the land guiding surfaces are Rulon. (Courtesy of ITI Corporation. From Reference 15.)

fluid or are lubricated by gear splash in much the same manner as oil-lubricated bearings (Figure 34). Another application is in the maintenance of suitable temperatures for the use of "superconductivity" electronic equipment. At a temperature of -452°F current will flow with very minute impressed voltages; cryogenic control is essential here. General Electric has developed a "cryo-gyro" that utilizes the increased effectiveness of magnetic fields at low temperatures to support a virtually frictionless gyrorotor.

Bearings for space use are designed for minimum heat generation and maximum heat dissipation by conduction. Designs must provide for thermal

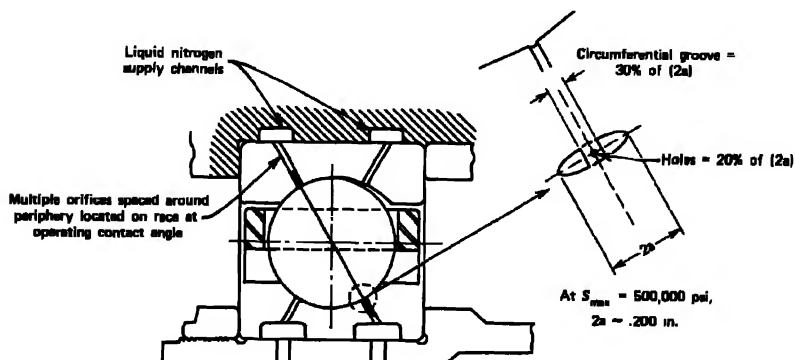


Figure 34. Cryogenic lubrication—a design to force the cryogenic fluid into the rolling contact zone. (Courtesy of ITI Corporation. From Reference 15.)

expansion. Minimizing bearing preloading is one technique of limiting heat generation. Another method is to use open-race curvature on races. Race curvature is defined as the ratio of the ball groove radius to the ball diameter. Conventional race curvature is about 51 to 52%. For better space characteristics, 54 to 58% is recommended. Since almost all heat generated must be rejected by conduction through the races, alloys providing good heat transfer properties must be considered. Although conventional materials such as 52100 and 440-C can be used at operating temperatures below 450°F, the danger of surface cold welding will be reduced if a cast cobalt alloy or a nonhomogeneous tool steel is used. Austenitic stainless steels and other homogeneous alloys should be avoided because of surface welding tendencies. Dimensional stability with thermal cycling is also an important design feature that prevents the use of austenitic stainless steels. At temperatures above 450°F, 52100 and 440-C steels suffer a loss in hardness that makes them susceptible to brinelling; tool steels are a good bet for this range.

For temperature ranges of 275 to 600°F iron silicon bronze is a good choice for a retainer material. At temperatures above 600°F copper alloys deteriorate and nickel alloys such as S-Monel should be considered.

8.4.5. Space Age Metallurgy

About 40 years ago the bearing industry standardized on 52100 steel for most bearing components. After World War II technological requirements forced a shift to vacuum processing for premium quality. Vacuum melting is a process where the cold charge is melted in an induction furnace and poured into ingots while the operation is under a vacuum. Although the finished product is superior to steel produced in the conventional manner, there is a wide variation in quality and the material is expensive. The consumable electrode vacuum process remelts electrodes made from a primary air-melted process by utilizing an electric arc. The liquid metal solidifies in a water-cooled copper mold under vacuum. This technique consistently produces high-quality steel but at a price five times greater than that of air-melted steel. The practical solution to the problem is called vacuum processing; this process utilizes steel melted in an electric air furnace that is subsequently treated in a vacuum. The result is steel quality approaching the level of vacuum-melted steel but at a moderate price. Three methods are presently used for vacuum processing:

1. Stream degassing.
2. Ladle degassing.
3. The Dortmund-Horder process (DH).

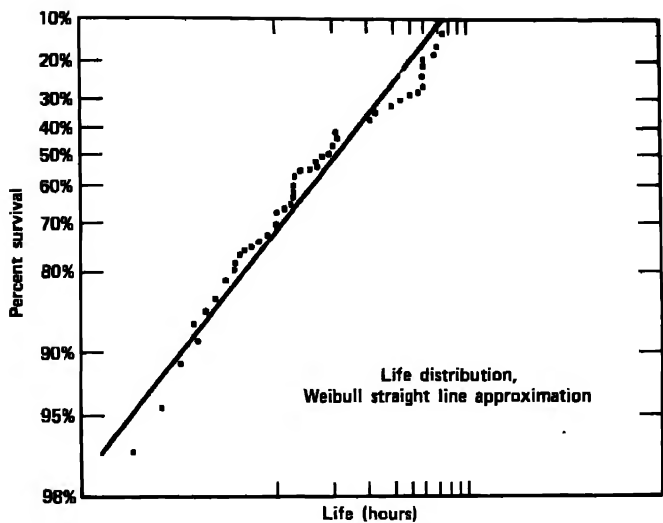
All of these processes involve exposure of the air-melted steel to a vacuum while it is in the molten state; this results in significant reduction in gas content compared to normal air-melt techniques. The inclusions found in normal bearing materials have a significant effect on bearing endurance. The vacuum process is particularly effective in eliminating oxides, silicates, and aluminates, which are the most injurious to steel. Air-melt processes normally use silicon and aluminum as deoxidizers; vacuum techniques utilize carbon, which produces gaseous by-products that leave no harmful residue. Another advantage of vacuum technology is that it permits close control of chemical analysis of steel, which is impractical with other techniques.

A method of evaluating the effectiveness of the endurance of vacuum process metals is provided by the Weibull plot (Figure 35a), which is an accepted method of displaying the life distribution of bearings. The abscissa is the duration of the test in time (usually hours or days) plotted on a logarithmic scale. The ordinate is the percentage of bearings surviving the test; the scale is plotted as $\log (1/S)$, where S is the percentage of survival. Standard bearings normally approximate a straight line having a slope of about 1.3. Better bearings have the line displaced to the right and have a slope greater than 1.3. Figure 35b shows the results obtained with NDUR, a product of the New Departure Division of General Motors. This material is produced by vacuum processing and is forged to shape. The B-10 life referred to is the life expectancy of 90% of the sample. This index clearly indicates the superiority of vacuum process over conventional steels. The significance of using vacuum-type metals is that the B-10 life is 3 to 6 times as great as ordinary steel, based on a 90% criterion rather than average life.

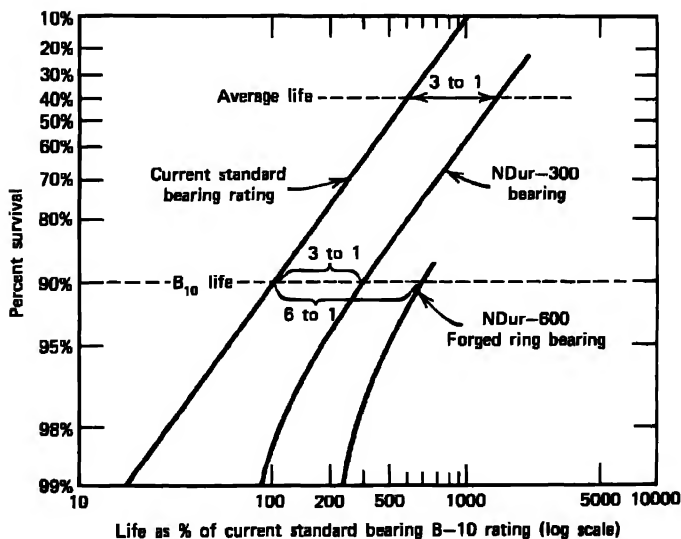
The development of bearings to fit beryllium components has also been accelerated by the space program. Beryllium is an excellent space metal for several reasons:

1. It has superior strength-to-weight ratio— $1\frac{1}{2}$ to 2 times that of aluminum.
2. Its strength-to-weight superiority extends to temperatures in excess of 1000°F.
3. It has a very high modulus of elasticity (40×10^6 psi) and excellent dimensional stability.
4. Its coefficient of expansion is almost identical to that of steel.

The problem of utilizing steel bearings in beryllium parts was first encountered in gyros where the float is made of beryllium. Another application is in telescope mounts where, for example, the three pads on which the telescope is mounted are lapped flat to within 10×10^{-6} in. and are also held coplanar within the same tolerance. The Barden Corporation has



(a)



(b)

Figure 35. (a) A typical Weibull plot. (b) Standard bearing life rating versus NDur-300 and NDur-600 ratings. (Courtesy of Society of Automotive Engineers.)

successfully built both types of bearings. The mating of steel bearings and beryllium components has two prime problems:

1. The relative brittleness of beryllium. Beryllium components must always be designed to be kept in tension during usage.
2. The avoidance of contamination of the material during processing. It is typically clad with a layer of mild steel for protection.

When mating steel and beryllium components, all large stresses must be avoided because of the poor ductility of beryllium. The parts are often chemically etched or finely polished to eliminate flaws incurred during machining. Additional limitations on beryllium include high cost and toxicity. Nevertheless, bearings for beryllium parts are an inherent space-age reality and the United States industry is now successfully meeting the challenge.

A basic criterion for the state of the art of bearings can be found in the specification being used today. A universally accepted standard has been published for many years by the Anti-Friction Bearing Manufacturers Association (AFMBA Standards). The least accurate bearings have been designated as ABEC -1; the higher the quality, the greater the ABEC number. Automobiles, for example, typically use ABEC 1 and 3. When the space era began, ABEC-5 bearings were used as the finest available product for sophisticated equipment. Today ABEC 7 and 9 are considered first-grade bearings, with many special applications exceeding these standards. To illustrate the basic difference between ABEC-7 and ABEC-9 bearing: the radial runout on a $\frac{3}{8}$ in. diameter ABEC-7 is 1×10^{-4} in., while for an ABEC-9 it is one-half that amount. Most tolerances for small bearings are now expressed in "tenths" (1×10^{-4} in.) rather than thousandths. Inspection technique problems now assume proportions almost as great as fabricating the bearing.

8.4.6. Commercial Design Trends

Although the aerospace industry has added tremendous impetus to basic bearing research, the vast majority of bearings are still used by nondefense industries. Here there are rarely state-of-the-art problems to contend with, but rather questions of a systemic refinement of well-known techniques. One of the chronic concerns has been the improvement of large, thin-section bearings. These components are typically used in rotating electrical equipment to provide rotation with minimum torque and maximum concentricity of mating parts. The contributing factors to bearing torque are well-known:

1. Friction between rolling elements and races.
2. Ball-to-ball friction.

3. Viscous friction due to lubricants.
4. Dynamic effects.

A more subtle but equally important contribution is made by bearing runout, or eccentricity. The force generated by a moving bearing with an eccentric component can be expressed as

$$F = (2\pi)^2 f^2 MA \quad (5)$$

where F = shaking force developed (pounds)

f = frequency of vibrations (revolutions per second)

M = mass of the eccentric part

A = one-half the peak-to-peak eccentricity

The eccentric member is typically the inner or outer race of a bearing and the load it supports; this may be a journal, a turbine rotor, or a gyro-wheel. The runout between the inner and outer races of bearings is therefore the most serious single defect in a bearing and may result in the generation of tremendous shaking forces. In addition to runout on races, single defects on the outer and inner race also cause shaking forces. A given defect in a race may lead to a shaking force every time a ball strikes it. A variation in the size of the ball has a similar effect. Although the amplitudes of these disturbances may be measured in microinches, the frequency of occurrence is often high and the effect appreciable.

Consequently, thin-section bearings 10 in. in diameter are now being produced, with radial runout tolerances of 0.0002 to 0.0003 in. and balls spherical within 5 μ in. Typical small R4-type bearings were no better a few years ago. The usual construction of the bearing includes stainless steel races, double shields, and spring separators to minimize torque. Coefficients of friction as low as 0.0004 have been reported.

Another interesting approach to reducing vibration and noise is provided by Fafnir Bearing Company (Figure 36). The basic bearing components are assembled into a Buna N rubber shell, which in turn is bonded into a metal outer shell. The Buna N shell has excellent damping qualities and can also be utilized at moderately high temperature. For higher temperatures (450°F) Viton is used. Typical applications decrease noise levels by 50% and also reduce axial and radial shaking forces that may be transmitted to the rest of the machine.

An example of good power transmission is provided by the Kaydon integral gear race. By incorporating gear teeth on a bearing race, the number of components is at least halved and space and weight are saved; above all, accuracy is far better (Figure 37). In applications using a separate gear and bearing, the total runout of the system comes from three sources: the runout

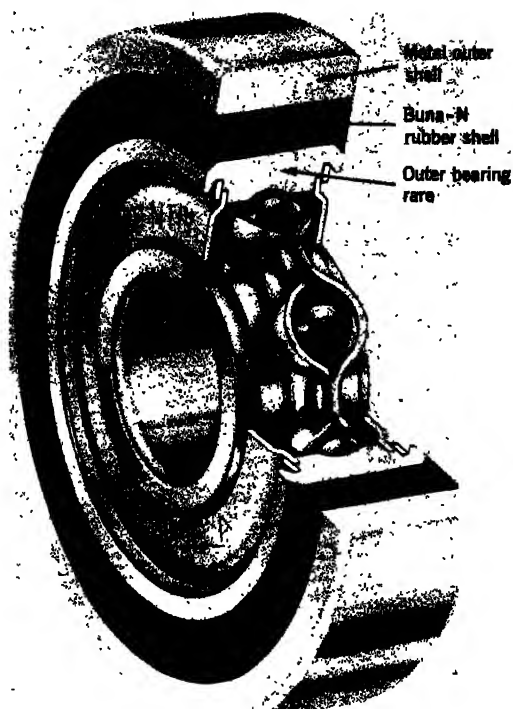


Figure 36. Quiet bearing construction. (Courtesy of Fafnir Bearing Company.)

in grinding the bearing races; the runout between the gear teeth and its piloting diameter; and eccentricity in mating the gear and bearing centerlines. By cutting the gear and grinding the bearing on a common ring, runout of the unit is confined to that normally found in only the bearing itself. Cutting the gear on the bearing race results in greater than normal gear strength since the bearing material is normally made of more costly alloys than gear blank materials. Where specifications call for a gear pitch that might cut into the hardened portion of the bearing race, a heavier race section may be provided. In general, the races and bearing loads are machined first and then used as piloting diameters for cutting the gears. Thus extremely close runout specifications can be maintained. Integral gear-bearing designs are used in radar antenna mounts, azimuth bearings, camera mounts, and cranes.

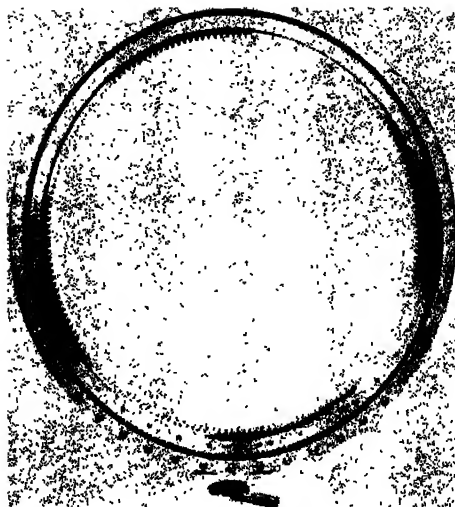


Figure 37. Bearing with integral gear race. (Courtesy of Kaydon Corporation.)

An unusual type of bearing designed to support linear-motion mechanisms is called the **roundway bearing**. It is designed to supplement the rotating ball bushing sleeve where high loads are encountered. The roundway bearing is a recirculating linear-motion roller bearing designed for use on hardened and ground circular ways (Figure 38). It has three basic parts:

1. Bearing race.
2. Roller assembly.
3. Eccentric trunion pin.

The bearing is normally fastened to the underside of a carriage by mounting block (4), rolls freely on a train of concave rollers (2), which bear on hardened circular ways (6). These roundways may be supported on the underside (7) to prevent deflection under heavy loads. By rotating the eccentric trunion pin (3), the height of the bearing can be adjusted to compensate for variations in the surface of the underside of the carriage and to eliminate buildup of tolerances between component elements. The "dot" on the end of the trunion pin indicates the high point of the eccentric. After adjustment, the height is "locked in" by simply tightening the lock-screw (5). The "rockable" bearing race, which is mounted on the trunion pin, guarantees uniform load distribution on the rollers. Because of this,

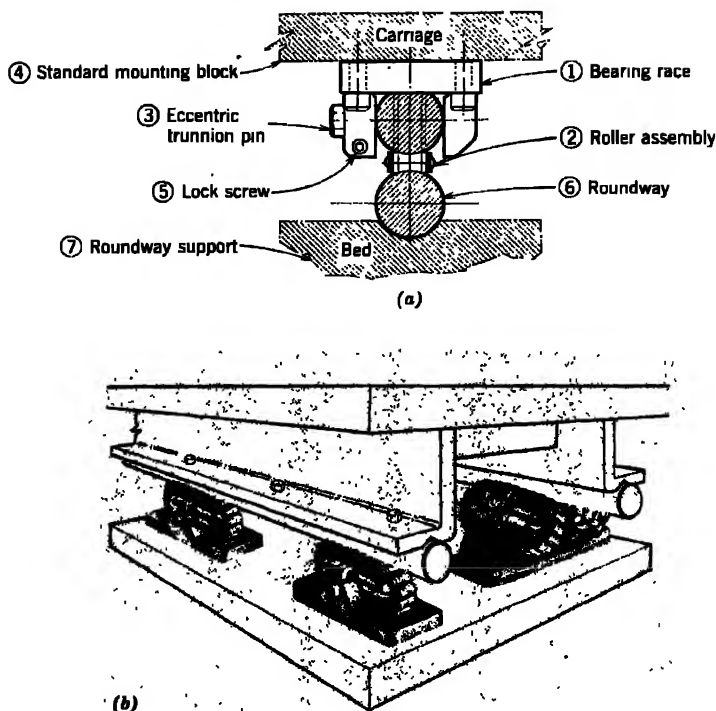


Figure 38. Roundway bearings. (a) Details of construction. (b) A typical installation. (Courtesy of Thomson Corporation.)

and the cylindrical slope, the bearing unit will self-align to some extent in all directions. The advantages of the roundway bearing are:

1. Greater load capacity than in linear bearings.
2. Low cost.
3. Elimination of high friction due to skewing found in roller designs.
4. Reduced wear compared to flat-way designs.
5. Convenient height adjustment and minimal end play.

Preloading is done by means of the eccentric trunnion pins. The unit can be sealed completely by standard enclosures available from the manufacturer.

Hollow bearings have the design advantages that at high speeds the centrifugal force exerted on the supporting structures is lower, have less effect on the contact angle of the bearing, and extends the fatigue life of the bearing.

Hollow balls are commonly made with walls of 0.005 to 0.020 in. thick. Conventional balls are fabricated by upset forging of wire slugs to form the initial spherical blank. Hollow balls are manufactured by joining together two hemispherically formed shells. In this process the grain flow within each hemispherical shell is completely circumferential and the end grains do not extend to the outer surface of the ball. The two halves are typically welded together with an extremely narrow "weld line." Test results indicate no fatigue or structural problems at these "seams." A conventional polishing operation follows. The deflection characteristics of the hollow spheres are substantially the same as of solid spheres (Figures 39 and 40). The three-dimensional characteristics of the ball serve to restrain its center portion and eliminate significant deflection due to "bending" of this ball.

Bearings utilizing hollow rollers as well as hollow balls are made by the Industrial Tectonics Corporation. See Appendix C for a glossary of bearing terminology.

8.5. POWER SUPPLIES

The most fundamental component in any electrical system is the power supply. It is a buffer between the electronic system elements and power sources that vary in frequency, voltage, and waveform. The majority of

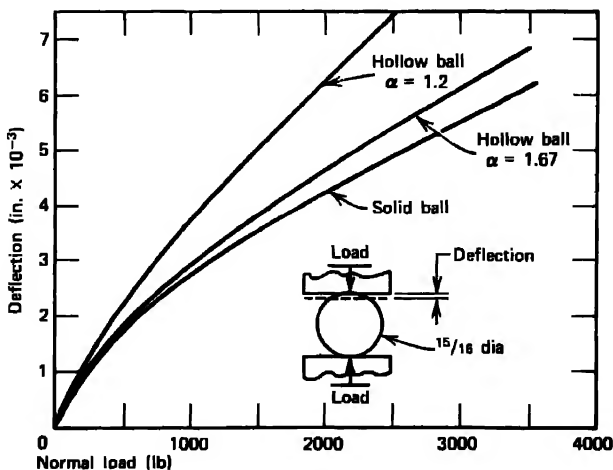


Figure 39. Deflection characteristics of hollow and solid balls. (Courtesy of ITI Corporation. From Reference 15.)

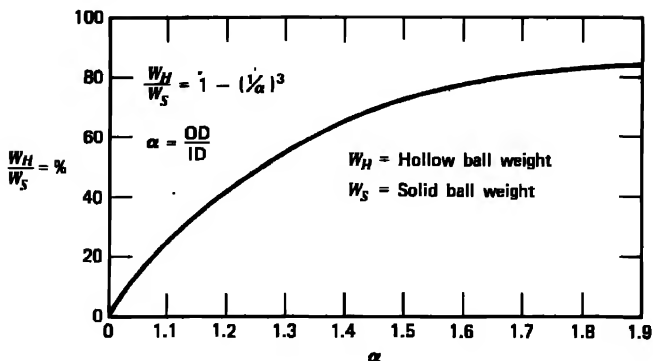


Figure 40. Hollow ball bearing characteristics. (Courtesy of ITI Corporation. From Reference 15.)

power supplies used in systems convert unregulated AC voltage to regulated DC voltage.

Power supplies are grouped in six primary categories:

1. *High-current DC.* Current output is above 5 A at 0 to 450 V. Input power is 115 V AC, 60 Hz.
2. *Constant-current DC.* The current output is 0 to 125 A at 0 to 2500 V. This category differs from high-current units mainly in that it regulates current while the latter regulates voltage.*
3. *Laboratory DC.* This category provides constant voltage at outputs up to 3 A. Regulation is precise.
4. *High-Voltage DC.* This group provides regulated output voltages up to 1,000,000 V with inputs of 115 V AC, 60 Hz.
5. *Special-Purpose DC.* This is a catchall category designed to bracket any power supply that does not fit in the first four categories. Typical examples are microwave supplies, voltage references, and extra precise units.
6. *Regulated AC.* This includes AC supplies with the following characteristics:

- (a) Amplitude-regulated.
- (b) Fixed-frequency.
- (c) Adjustable-frequency.

* A good constant-current power supply is a poor constant-voltage source and vice versa. When both constant current and constant voltage are required, separate sections of the power supply must be provided.

Some literature and manufacturers' catalogs also list a seventh category—modular power supplies. These are compact power units that are designed to plug into certain classes of chassis. They are completely self-contained and are available in convenient ratings that are physically interchangeable. All six categories have modular configurations.

8.5.1. Basic Configurations

Modern power supplies are derived from a few basic circuits with countless variations and combinations to produce special performance characteristics.

The series regulator circuit shown in Figure 41 is an example of a classic design that is still widely used. The input AC voltage is supplied to a step-down transformer and then to a ring-type demodulator (4 diodes) that

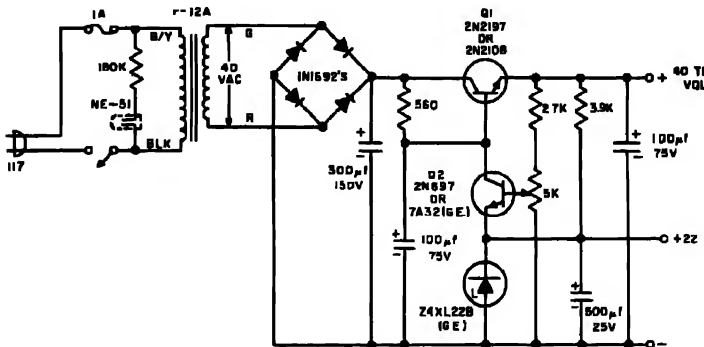


Figure 41. Regulated DC voltage supply. (Courtesy of General Electric Company.)

converts it to pulsating DC voltage. A large capacitor smoothes out the pulsation. The signal is then fed to series regulator Q_1 which acts as a variable resistor in series with the load. Transistor Q_1 is controlled by error sensor Q_2 which regulates the base current to Q_1 . The emitter of Q_2 is held at a fixed voltage by zener diode Z_4 ; the base senses the output voltage of Q_1 . When the output voltage increases, base of Q_2 turns on the transistor, thus lowering the base voltage of Q_1 ; this decreases the current output from Q_1 , which in turn lowers the voltage at the base of Q_2 , and the cycle continues. The outer capacitors shown in the circuit are used for eliminating transients from the output.

A more sophisticated series regulator is illustrated in Figure 42. At this point it should be noted that most literature on power supplies omits the

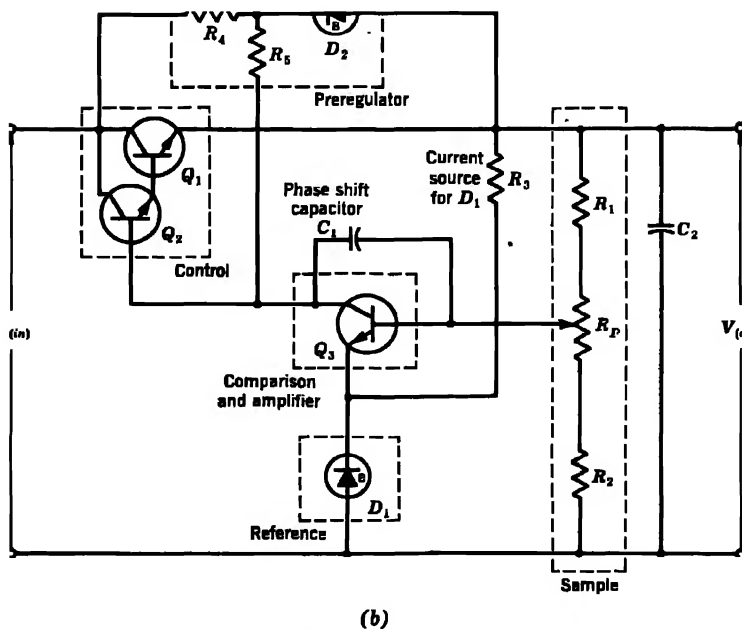
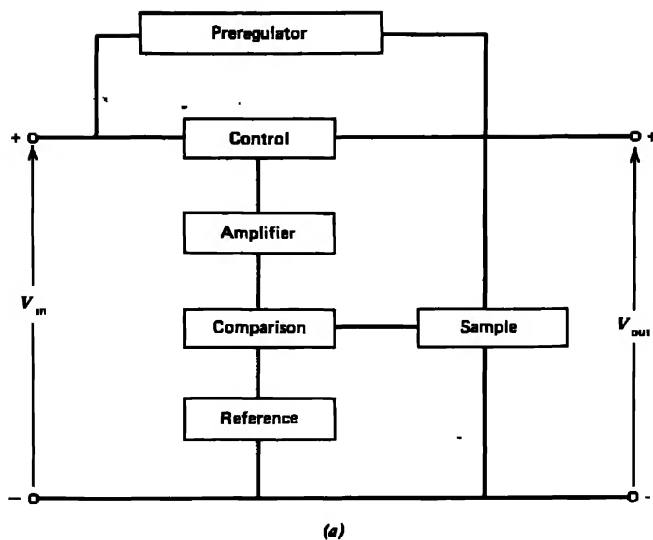


Figure 42. Series voltage regulator. (a) Block diagram of a series or emitter-follower DC voltage regulator. (b) Schematic diagram of a series DC voltage regulator. (Courtesy of Texas Instrument Inc. From Reference 32.)

rectifying elements in the power supply—the input transformer, ring rectifiers, and smoothing capacitors. This is done because some of the circuits shown can also be used with DC inputs.

The circuit of Figure 42 has six principal elements:

1. Reference circuit.
2. Comparison circuit.
3. Amplifier circuit.
4. Control circuit.
5. Preregulator circuit.
6. Sample circuit.

The heart of this power supply is the reference element; it is usually a zener diode and its associated current-source resistor. This provides the "standard" voltage for all other circuits in the power supply.

The comparison circuit senses the difference in voltage between the reference element and the output and varies the voltage on the base of the control circuit. As the base voltage on Q_3 (Figure 42*b*) increases, it draws more collector current, which in turn lowers the collector voltage because of the increased voltage drop in resistor R_3 . This also lowers the base voltage on Q_2 , which decreases the control circuit output voltage. The capacitor C_1 is used to correct the phase shift between the base and collector of Q_3 .

The amplifier circuit is utilized to enlarge the difference in voltage sensed by the comparison circuit. The circuit shown in Figure 42*b* does not contain a separate amplifier circuit. Another common-emitter stage is usually used. When the requirements of the amplifier are very high, a difference type of amplifier is needed.

The control circuit is essentially a solid-state "variable resistor" that regulates the current flow to the load in accordance with voltages received from the comparison and amplifier circuits. It consists of one or more transistors with their collectors connected to form a beta multiplier circuit. The last stage has its base wired to the amplifier circuit so that the entire control circuit is capable of amplifying the error signal received from the comparison circuit.

The preregulator supplies a constant source of current to the amplifier—comparison circuit and the base of the control circuit. This helps to prevent ripple from reaching the control circuit where it would be further amplified.

The sample circuit is simply a resistor and potentiometer network that provides the comparison circuit with a properly scaled voltage.

Shunt regulators use a device in parallel with the output that maintains a constant voltage as the current varies. The components are typically zener diodes, gas tubes, or vacuum tubes (Figure 43). The majority of shunt regulators on the market today are considered to be reference voltages

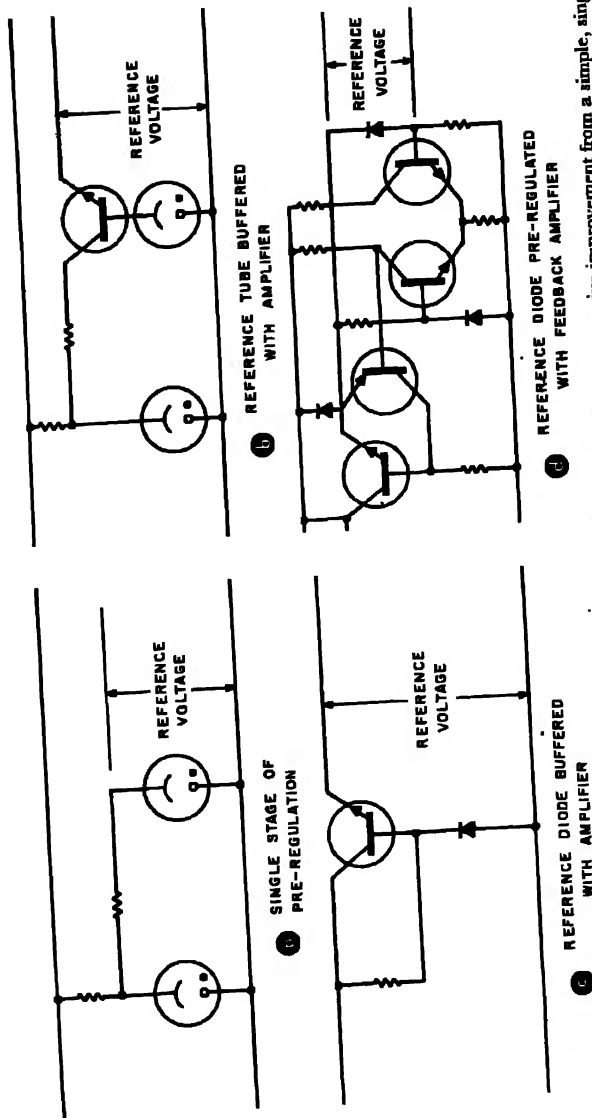


Figure 43. Shunt voltage regulator. The reference voltage circuits have undergone progressive improvement from a simple, single stage of preregulation (a) to a reference diode preregulated with a feedback amplifier (d'), which is actually a full-scale power supply. (Courtesy of Hayden Book Company. From Reference 21.)

rather than power supplies, since they are best suited for low-power applications. Vacuum-tube shunt regulators are used in low-current, high-voltage power supplies. Shunt regulators are basically less efficient than series regulators; consequently they are employed largely as auxiliary components in other types of power supplies.

A third generic type of power supply is the SCR (silicon-controlled rectifier). The phase control of a pair of SCRs is used to control output voltage (Figure 44). In a typical circuit, the firing point of the SCR pair,

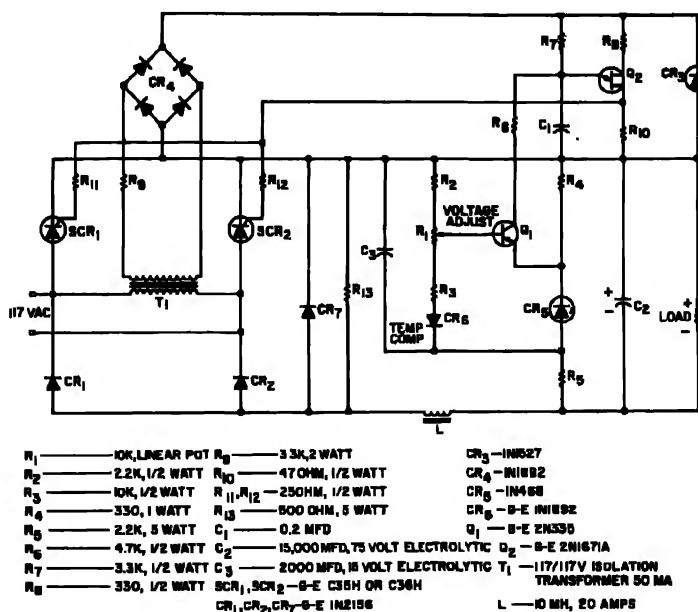


Figure 44. Silicon-controlled rectifier—phase-controlled constant voltage power supply. (Courtesy of General Electric Company. From Reference 22.)

SCR₁ and SCR₂, is controlled by the unijunction transistor Q₂. A conventional ring rectifier, CR₄, provides the DC power, and RC network R₇C₁ provides the required pulse-shaping network. The regulating action is achieved by comparing the base voltage of Q₁, which is proportional to the output voltage, with the reference voltage across reference diode CR₅. If the load voltage tries to rise, more base current flows through Q₁. The resultant increase in collector current in Q₁ diverts charging current from C₁ and lengthens the time needed to reach the peak voltage required to

fire Q_2 . This retards the firing angle and returns the load voltage to normal. If the load voltage drops, the reverse action takes place. The two SCRs and two diodes, CR_1 and CR_2 , form a bridge with the load on one end. The longer the SCRs are in the conducting state, the higher the output DC voltage.

One of the most significant developments is the bridge circuit power supply. It is an arrangement of previously discussed components to form a conventional "diamond-shaped" bridge (Figure 45). The voltage delivered

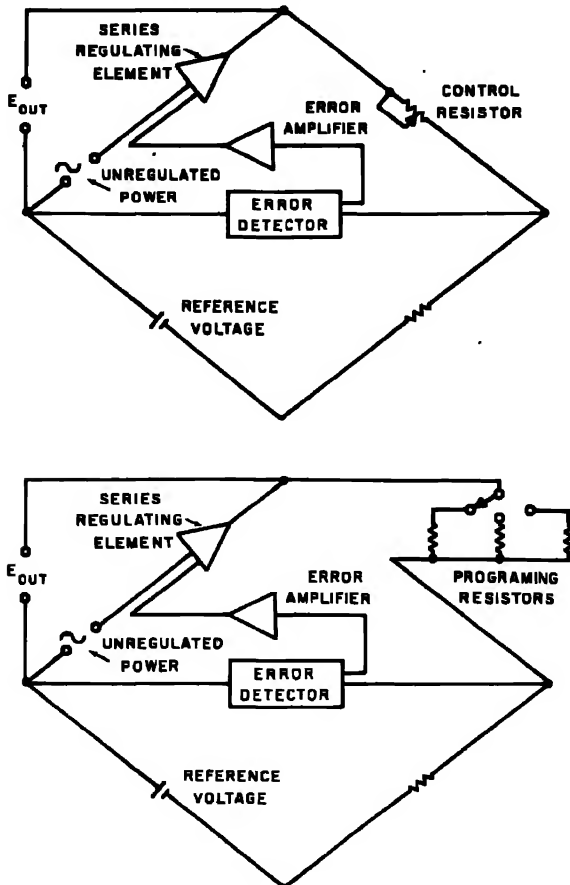


Figure 45. Bridge-type series regulated supplies were a major development in the progression of power supplies. This made possible the programming of power supply output. (Courtesy of Hayden Book Company. From Reference 21.)

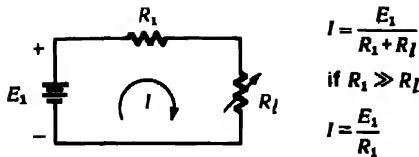


Figure 46. Constant current power supplies. Series resistance method of simulating a current source. (Courtesy of Kepco, Inc.)

to the load is a function of the control resistor. If this element is changed, the power supply can be used to deliver more than one controlled voltage. In addition, it is possible to arrange a switching network so that the power supply can be programmed to deliver a specific voltage at a fixed time or in a given cycle. So far all of the circuits described have been of the constant-voltage type. To construct a constant-current power supply, it is necessary to make the load resistance, R_L , much smaller than the resistance of the regulator (Figure 46). This can be achieved by using the circuit shown in Figure 47, where E_r is a reference voltage, R_r its associated current-limiting resistor, E_u the unregulated source of DC power, and R_{uc} its associated feedback resistor. The reference and unregulated voltages are normally at "null" at the input to the amplifier. When E_u falls below E_r , an input signal is directed to the amplifier and an inverted output is directed from the amplifier to the base of the series control transistor. This causes the transistor to pass more current, increasing the voltage across R_{uc} , thereby bringing the system back to "null." The output current to E_o is held constant by this action. It is important to note that a very small current input to the amplifier controls a relatively large load current. This approximates the action of the transformer shown in Figure 47b.

The basic difference between constant-voltage and constant-current power supplies is shown in Figure 48. If the feedback is taken from a voltage-sensitive point, as in Figure 48a, the output voltage is regulated at a value equal to the reference multiplied by the ratio of the proportioning resistors.

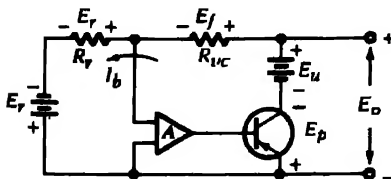


Figure 47. Null-type input power supply. (Courtesy of Kepco, Inc.)

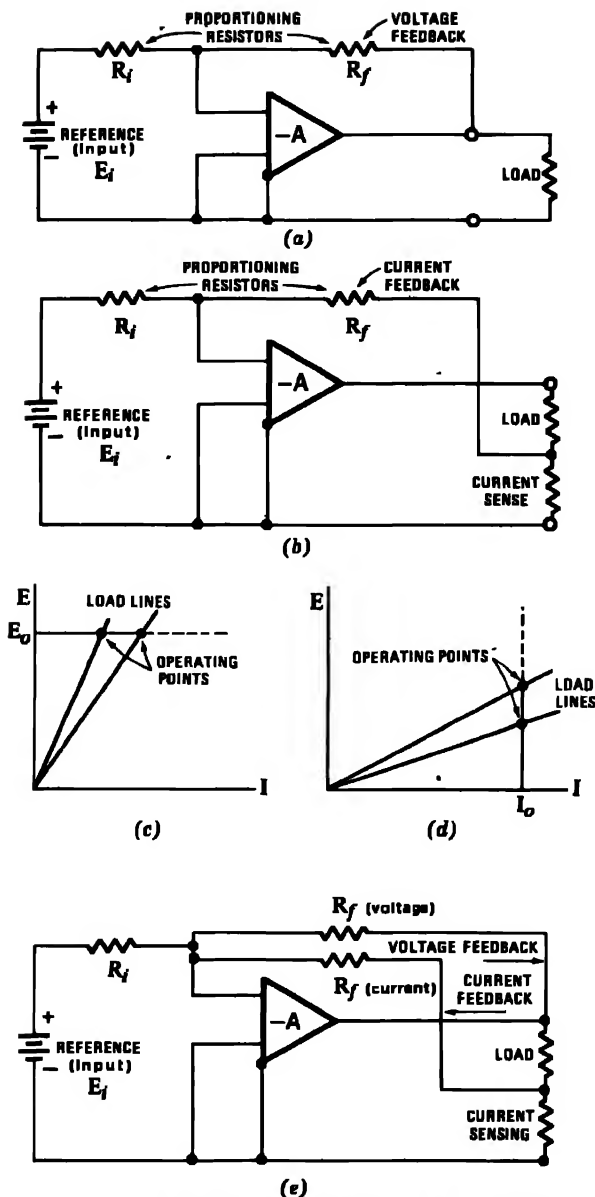


Figure 48. Current and voltage feedback circuits used in power supplies. (a) Voltage feedback. (b) Current feedback. (c) Voltage regulation. (d) Current regulation. (e) Simultaneous voltage and current feedback. (Courtesy of Kepco, Inc. From Reference 24.)

It may be varied by changing either the reference itself or the proportioning resistors. The device is termed a voltage regulator and has an E/I characteristic, as shown in Figure 48c.

Feedback from a current-sensing point, as in Figure 48b, results in the regulation of current with the complimentary output characteristic (Figure 48d).

8.5.2. Selection of Power Supplies

The two most commonly used power supplies today are the series regulator and the SCR type. Here we discuss some of the most important characteristics that must be considered when selecting a power supply, with emphasis on these two types. Again, it is important to realize that the recommendations given are subject to technological obsolescence. Moreover there are no "yes or no" answers to questions about power supply selection; that is why there is room for so many good companies in the business.

LOAD. The first consideration is to find a power supply that meets the system requirements for voltage and current, with a reasonable safety factor. The safety factor is expensive and should be selected carefully. SCR units are normally best suited for high-power output (1 kW or more), while series regulators are used for the low- to medium-power range (1 to 500 W).

LINE REGULATION. Line regulation is the change in output current or voltage for a given change in line voltage. Series regulators hold this value to 0.01 to 0.1%. SCR supplies run between 0.1 and 0.5%.

LOAD REGULATION. Load regulation is the change in output for a specified change in the load while the input voltage is held constant. Nominal tolerances are about the same as for line regulation.

STABILITY. Stability is a change in output without a change in line voltage or load. Normal performance figures are specified after an initial warm-up period. Series regulator circuits normally are held to 0.1% while SCR supplies are within 1%. Some companies specify not only the warm-up period, but also the time interval for the stability figure, such as a 15-min warm-up period and an 8-hour stability interval.

AMBIENT TEMPERATURE VARIATION. This is defined as the percentage change in the output voltage or current due to a variation of ambient temperature. The values are usually expressed as a percentage per degree centigrade over a prescribed ambient range of the unit. Typical figures are 0.015%/°C for series regulators and 0.03%/°C for SCR circuits in the range of -20 to +55°C.

RIPPLE. Ripple is that portion of the output voltage which is harmonically related in frequency to the input power and to any internally generated switching frequency. It is expressed as an RMS percentage of output voltage or as an absolute peak-to-peak voltage. The National Electrical Manufacturers Association (NEMA) anticipates that the term ripple will be replaced by PARD*, which will include hum and noise as well as spikes in the output.

RESPONSE TIME. This is the time required for a current or voltage transient caused by a load change to return to 37% of its maximum overshoot after a step load change. Typical solid-state series regulator units require less than 50 μ sec. Vacuum tube circuits are in the 30-msec range.

COST. After all the parameters listed above are thoroughly checked, the ultimate common denominator is price. SCR power supplies are among the least expensive. Series-regulated units are considerably higher-priced. Before submitting the purchase order, be certain to verify that the unit of your choice can operate successfully on the raw power available to the project. This includes variations in the nominal supply voltage and frequency.

Some of the other pertinent parameters are listed in the glossary given in Appendix D.

8.5.3. Options

In addition to the basic parameters discussed above, a number of options exist, which are sometimes very useful.

OVERCURRENT CONTROLS. Some type of fuse, magnetic circuit breaker, or current-limiting device is a must. The better power supplies have some form of protection that gradually reduces the current to a safe value instead of abruptly terminating all power.

THERMAL OVERLOAD CIRCUIT. This protects the power supply against current overloads, environmental overloads, or a defective, overheated component.

UNDERVOLTAGE ALARMS. A circuit is provided to sound an alarm when the voltage drops to a point at which it might cause system damage.

SURGE PROTECTION. Some power supplies have circuitry designed to reduce sudden surges that occur under abnormal circumstances. A zener diode circuit is one simple example.

* PARD is defined as periodic and random deviation of a DC output quantity from its average value over a specified bandwidth, with all influence and control quantities held constant. It can be expressed in RMS or peak-to-peak voltage values.

COOLING. A small fan is usually a valuable asset for high-duty power supplies.

SERIES/PARALLEL OPERATION. Some power supplies are designed to be used with similar units so that they automatically switch to series operation when additional voltage range is required and to parallel operation when more current is required.

PROGRAMMING. Automatic variation in voltage delivered is available according to a predetermined cycle.

AUTOMATIC CROSSOVER. Some power supplies have separate constant-voltage and constant-current bridge circuits. Automatic crossover is a feature that automatically switches the power supply to the appropriate circuit for a given power requirement.

DIGITAL CONTROL. If the control resistor shown in Figure 45 was replaced by a decade resistance box, the control of the output voltage could be adjusted minutely. This is exactly what is done in precision digital control units, except that the control box size is reduced to practical proportions.

8.5.4. Glossary

A glossary of the terminology used in power supplies is reproduced by courtesy of the Kepco Corporation, in Appendix D.

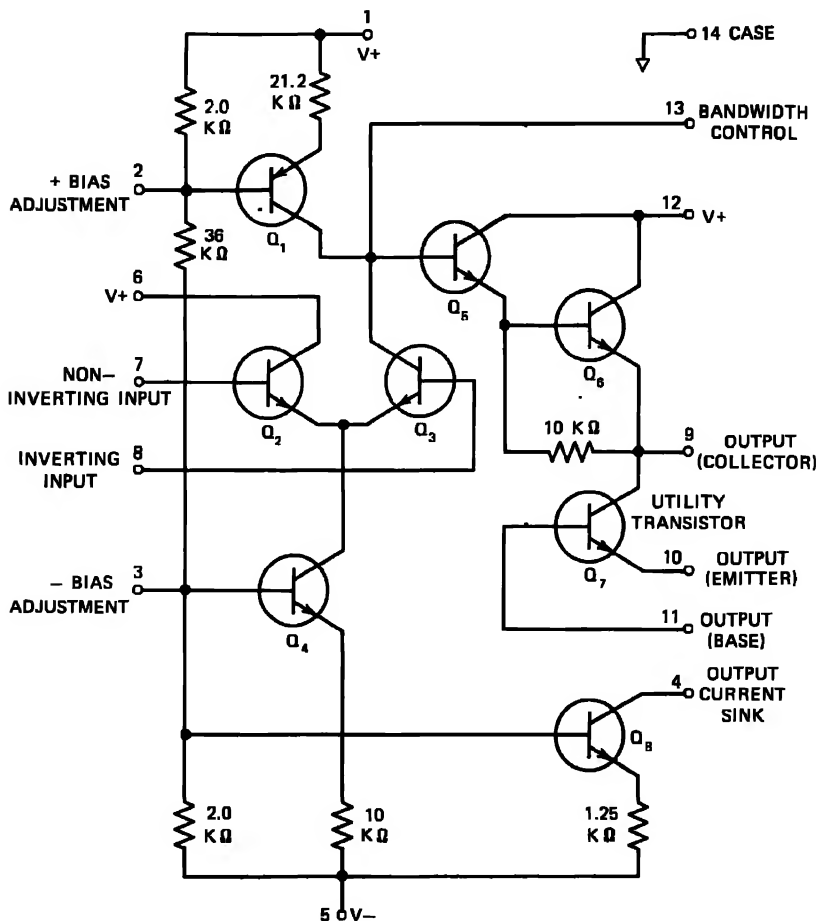
8.6. OPERATIONAL AMPLIFIERS

During the past ten years operational amplifiers have become a key element in system work. They are considered to be almost as basic as resistors, capacitors, and inductors. No other amplifier is equipped to do the wide variety of work routinely performed by this device. The name "operational" amplifier was used initially because the circuit was developed for mathematical applications; computers used it for such operations as addition, subtraction, multiplication, and division. The early units were designed with vacuum tubes; the second generation used solid-state components encapsulated in a small chip form. Current state-of-the-art utilizes integrated circuits. The last two categories now encompass about 95% of all units sold. Operational amplifiers have the advantages of high gain, low-drift rates, small size, low-power consumption, reliability, and low cost. They are used in computers for mathematical operations, integrators, filters, audio amplifiers, comparators, and other devices.

8.6.1. Circuitry

Operational amplifiers are designed to closely approach the ideal characteristics of any amplifier: high gain, high input impedance, low output impedance, large bandwidth, and very low internal noise. Figure 49 illustrates a typical operational amplifier.

The electrical path through Q_1 , Q_3 , and Q_4 forms a constant-current circuit. When a positive voltage is applied to the noninverting terminal,



COMPONENT VALUES SHOWN ARE TYPICAL

Figure 49. Schematic of a typical operational amplifier. (Courtesy of Radiation, Inc.)

7, the current through Q_2 increases, the current through Q_3 decreases, and the voltage at the collector of Q_1 increases. This in turn enlarges the current through the Darlington connection of Q_5 and Q_6 . Let us assume that the output terminal 9 is connected to terminal 4. For this case the increased current available at the emitter of Q_6 causes the voltage at terminal 9 to rise. This is the same polarity as the original input at terminal 7. When the positive input is applied at inverting terminal 8, the output at terminal 9 decreases, or inverts. This design permits the output to be obtained from terminals 9, 10, or 11, which correspond to the collector, emitter, and base of Q_7 . The choice depends on which impedance will be the most useful for the balance of the system. In each case the appropriate terminal must be connected to the current sink terminal 4.

Most operational amplifiers use modifications of this basic circuit to provide various advantages. The available circuits are very numerous and usually more complex.

Operational amplifiers are high-gain, DC-coupled amplifiers with either a differential or a single input. Signals applied to the noninverting terminal are amplified by a gain of A at the output. Signals applied to the inverting terminal are amplified by a gain of $-A$, or inverted. When only one terminal is used, the other is grounded. When employed in the differential mode, the output voltage is $A(E_2 - E_1)$ (Figures 50a and 50b). Single-input models are also available. The input and output impedances are represented as Z_{input} and Z_{output} , respectively.

Under ideal conditions, which are never fully realized, the design goals are as follows:

Gain approaches infinity.

Output voltage is zero when $E_1 = E_2$.

Input impedance approaches infinity.

Output impedance approaches zero.

Bandwidth approaches infinity.

Using these assumptions, refer to Figure 50b, which shows the amplifier isolated with gain A . Two external resistors are added: R_1 , an input resistor, and R_o , a feedback resistor.

Since the idealized amplifier input impedance is infinite, no current flows into the amplifier. The input current, I_i , is equal to the output current, since no current is diverted to the amplifier. The junction of the input and output current branches is commonly called the summing point:

$$I_i = \frac{E_i - E_s}{R} = \frac{E_s - E_o}{R} = I_o \quad (6)$$

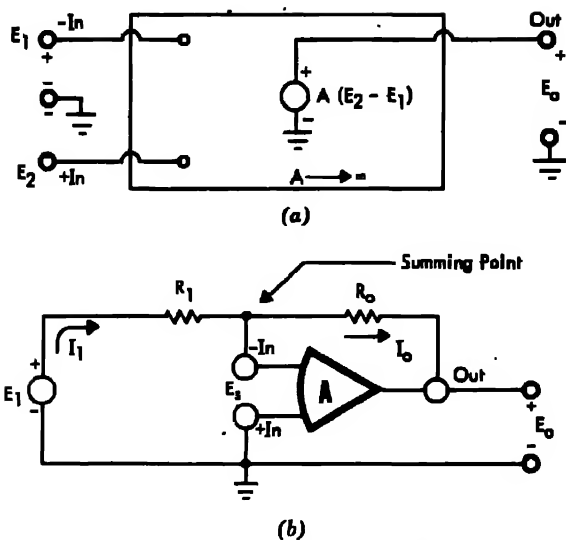


Figure 50. (a) Idealized operational amplifier model. (b) Inverting amplifier circuit. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)

However, $E_o = -AE$ due to amplifier action:

$$E_s = \frac{E_o}{-A} \quad (7)$$

Substituting equation 7 into equation 6 gives

$$I_i = \frac{E_i + (E_o/A)}{R_i} = \frac{(-E_o/A) - E_o}{R_o} = I_o \quad (8)$$

As A approaches very large numbers, or infinity, we have

$$\frac{E_i}{R_i} = \frac{-E_o}{R_o} \quad (9)$$

$$\frac{E_i}{E_o} = \frac{-R_i}{R_o} \quad (10)$$

Expression (10) is the transfer function for the closed loop gain of an operational amplifier and is well worth memorizing.

8.6.2. Typical Applications

SUMMING AMPLIFIER. The most common use of an operational amplifier is as a voltage amplifier. If the feedback resistor and input resistor are made equal, the gain of the circuit is unity. Under these conditions the operational amplifier can be used to sum a number of voltages (Figure 51). The advantage of this technique is that there is no interaction between inputs and the output voltage is independent of the load of the amplifier within its rated power level.

INTEGRATOR. When the feedback resistor is replaced next to a capacitor, the circuit becomes an accurate integrator (Figure 52). It is important to note that the integration process is limited by the saturation voltage of the operational amplifier. The switch S_1 is included symbolically as a means of discharging the capacitor. In practice, most integrators use some form of automatic device to perform this function and record the data on some type of permanent memory device. Other typical configurations are discussed in References 16, 17, 18, and 19.

8.6.3. Characteristics

Like all real components, operational amplifiers have certain limitations that must be understood.

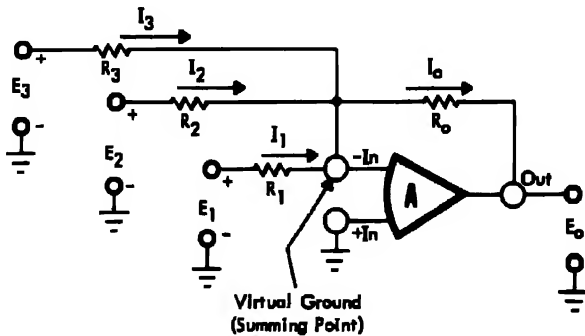


Figure 51. Summing amplifier is an extension of the fundamental inverter in which additional signal sources and summing resistors are added. Since the summing point is at virtual ground, the currents through the input resistors are $I_1 = E_1/R_1$, $I_2 = E_2/R_2$, and $I_3 = E_3/R_3$. All input signals flow through R_o , generating an output voltage $E_o = -I_o R_o = -(I_1 + I_2 + I_3) R_o = -E_1(R_o/R_1) - E_2(R_o/R_2) - E_3(R_o/R_3)$. If all three input resistors are equal, the circuit functions as a scaling adder. If differing values of input resistance are used, E_o is a weighted average of the inputs. (Courtesy of Burr-Brown Research Corporation. From Reference 28.)

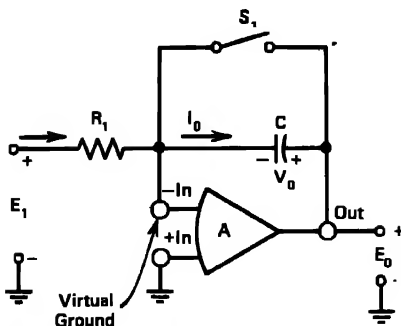


Figure 52. Integrator circuit. The integrator circuit has a capacitor as the feedback element. Current flowing into this capacitance is determined by the input circuit, $I_0 = E_1/R_1$. The amplifier output voltage is equal to the capacitor voltage, $E_0 = V_c = -1/C$

$$I_0 dt = -1/R_1 C$$

$E_1 dt$. Thus the output voltage is equal to the time integral of the input signal voltage, E_1 , with the scale factor, $1/R_1 C_1$. Another way of arriving at the same conclusions is to substitute $Z_0 = 1/C_s$, $Z_1 = R_1$ into the equation $E_0 = -(Z_0/Z_1)E_1$. This gives $E_0 = -(1/R_1 C_s)E_1$ or $E_0/E_1 = -1/R_1 C_s$, which is the transfer function of an integrator. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 28.)

Voltage and current offsets are parasitic effects that are generated within the amplifier. They cause a minute change of output with time and temperature. Input voltage offset is due to unequal base-emitter voltages in the input pair of transistors in the operational amplifier. Although all manufacturers make every effort to minimize this effect by matching techniques, the voltage offset is present and measurable. Variations in the voltage offset due to temperature and power supply fluctuations are called drift. Typical voltage offsets are shown in Figure 53a. Voltage drift is specified as the average slope of offset voltage versus the applicable parameter ($\mu V/^\circ C$). It is important to note that the definition of average voltage drift makes it impossible to distinguish between the two curves shown in Figure 53a. In Figure 53b we see that the voltage offset at the output is multiplied by $1/\beta$, where β is the feedback ratio.

Current offsets are also caused by the leakage or bias current in each input terminal of the amplifier. The current flowing into each input terminal is called the input bias current. As with the voltage offset, the current bias varies with temperature and power supply fluctuations. Figure 54 illustrates this problem. An equivalent input offset voltage arises as the product of the difference of bias current and R_1 . External circuitry can be used to substantially correct this effect. See Figure 55 and References 17, 18, and 19.

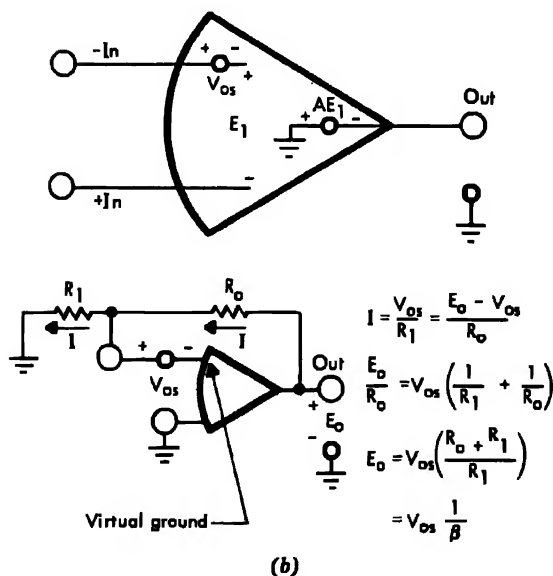
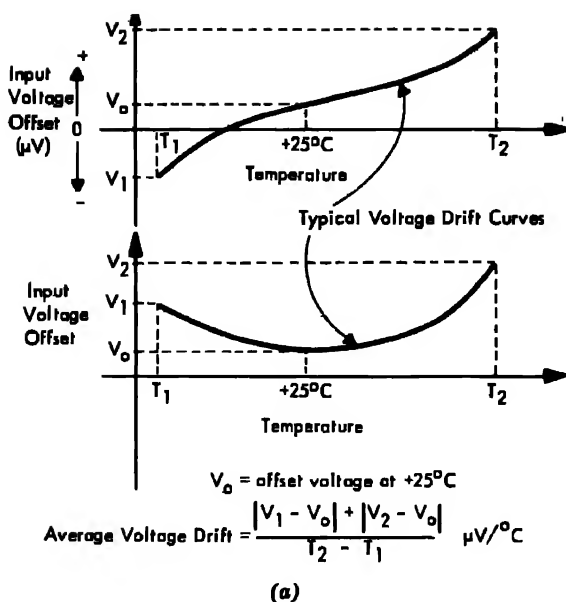
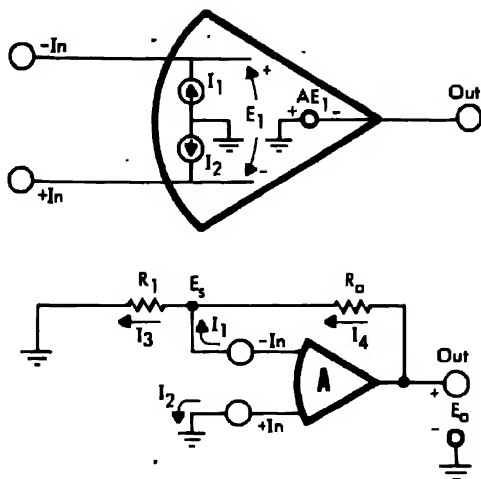


Figure 53. (a) Illustration of voltage offset and drift versus temperature. (b) Voltage offset model and the effect of voltage offset. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)



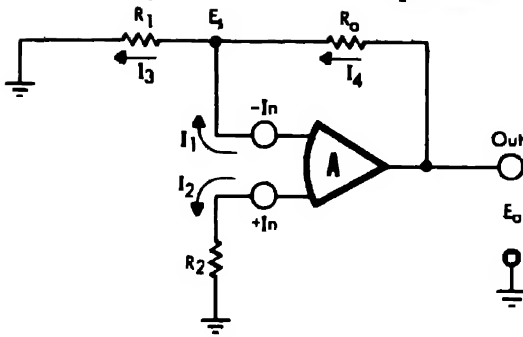
$$I_3 = I_1 + I_4 = E_3/R_1 \text{ and } I_4 = \frac{E_o - E_3}{R_o}$$

if $A \rightarrow \infty$, $E_3 \rightarrow 0$ then $I_1 \approx -I_4$

$$E_o \approx -I_1 R_o = \underbrace{(-I_1 R_1)}_{\text{equivalent input voltage offset}} \times \underbrace{\frac{R_o}{R_1}}_{\text{closed loop gain}}$$

Figure 54. Bias current model and the effect of bias current. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)

Drift can be defined as the change in offset current and voltage with time. Long-term drift may average out to zero, but short-term drift is often random. The causes have been traced largely to external effects such as minute changes in temperature, humidity, and power supply that affect basic transistor parameters. Since temperature has such an important influence on semiconductors, drift is normally specified over a definite temperature range. Drift with time at a fixed temperature is also specified. As transistor quality-control procedures improve and circuitry advances, the problems associated with drift will decrease. The most effective way of minimizing drift is to use a supplementary chopper circuit in the DC path of the operational amplifier (Figure 56). The chopper effectively keeps the



$$I_3 = I_1 + I_4, \quad E_1 = I_2 R_2 \text{ and } I_3 = \frac{E_1}{R_1} \text{ then}$$

$$I_3 = I_2 (R_2 / R_1)$$

$$I_4 = \frac{E_o - E_1}{R_o} = \frac{E_o}{R_o} - \frac{I_2 R_2}{R_o}$$

$$I_1 = I_3 - I_4 = \frac{I_2 R_2}{R_1} - \frac{E_o}{R_o} + \frac{I_2 R_2}{R_o}$$

$$E_o = -R_o I_1 + \left(\frac{R_o + R_1}{R_1} \right) R_2 I_2$$

$$\text{If } R_2 = \frac{R_o R_1}{R_o + R_1} \text{ then } E_o = R_o \overbrace{(I_2 - I_1)}^{\text{differential offset current}}$$

Figure 55. Use of compensator resistor R_2 to reduce the effect of bias current. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)

potential at the negative terminal of the operational amplifier at zero as current and voltage offsets are developed in the system; this practically eliminates drift. The circuitry functions as follows. The input voltage at E_1 is filtered at the summing point by the $C_1 R_1$ network into AC and DC components. The AC component is directed toward the operational amplifier, and the slowly varying DC component is directed toward the chopper. Network $R_2 C_2$ controls the frequency response of the input signal to the operational amplifier. The chopper converts the slowly varying DC component into an AC frequency that can be handled efficiently by the stabilizing amplifier, which typically has a gain of 1000. The amplified signal is then demodulated and further filtered by $R_3 C_3$. It is then fed to the positive terminal of the operational amplifier to correct the drift voltage originally

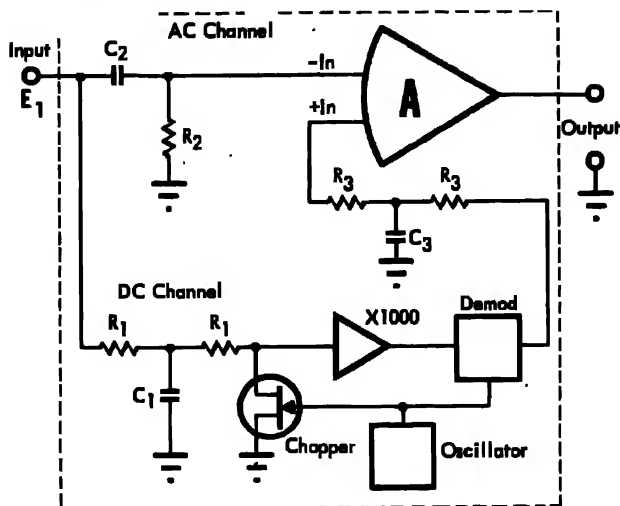


Figure 56. Chopper stabilized operational amplifier. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)

sensed at the summing point. Some chopper-stabilized amplifiers use a reference voltage instead of ground potential to produce more refined results. The overall circuit is designed so that the low-frequency roll-off of the AC path complements the high-frequency roll-off of the DC path. This technique reduces overall drift by a factor of 100 to 1000. Originally choppers were electromechanical devices, but modern units are solid-state components. A typical frequency response curve is shown in Figure 57.

Noise bandwidth is another major limitation on operational amplifier performance. Noise is a catchall term that includes all unwanted signals not related to input signals. Mathematically it is expressed as follows:

$$f_b = \frac{1}{G_0} \int_0^{\infty} |G(j\omega)| df \quad (11)$$

where f_b = effective noise bandwidth

G_0 = amplifier power gain at DC

$G(j\omega)$ = magnitude of amplifier power gain as a function of frequency

df = increment of frequency

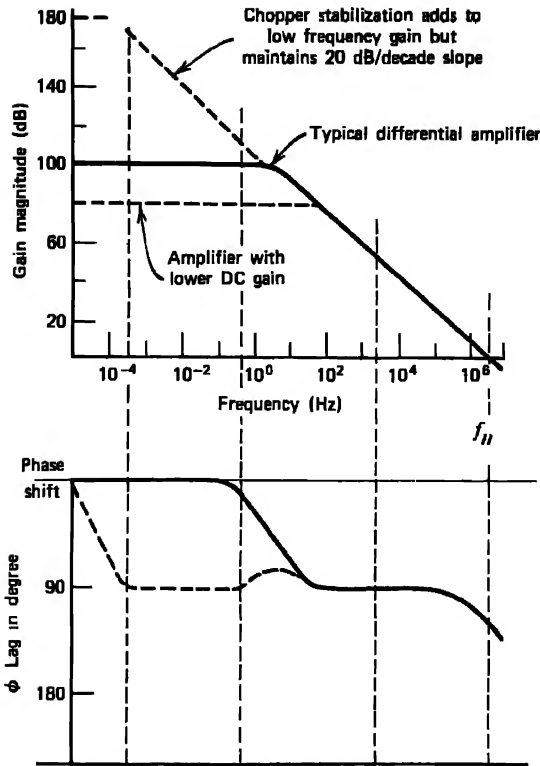


Figure 57. Open-loop gain and phase shift versus frequency for a chopper stabilized amplifier. (Courtesy of Burr-Brown Research Corporation, Inc.)

In practice, noise is arbitrarily separated into three bands:

1. From 0.1 to 10 Hz—generated by structural imperfections in the circuitry.
2. From 10 to 1000 Hz—usually traceable to pickup from the power line or chopper. It also includes associated harmonics.
3. From 1 to 100 kHz—a broad-band spectrum that is generated by resistor (Johnson or thermal) and shot noise.

Several methods are used to describe noise in technical literature. One way is shown in Figure 58. This simply provides specific RMS noise voltage for a given test technique. Another frequently used term is the noise factor, which is defined as the ratio of the total noise power in that band, as measured

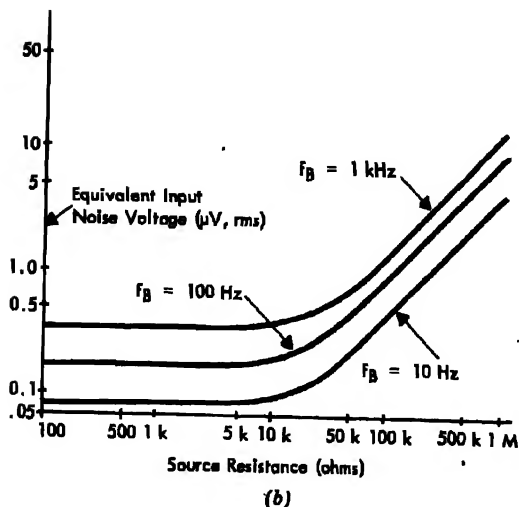
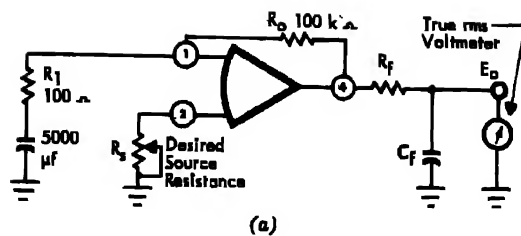


Figure 58. (a) Amplifier noise test circuit. The 5000- μ F capacitor gives a very low cutoff frequency, which may be considered to be DC for most measurements. Filter components C_f and R_f are selected to attenuate the output signal -3 dB at the desired bandwidth with further attenuation at 6 dB per octave beyond this point. Dividing this output voltage by the circuit gain E_o (in this case 1000) gives the equivalent noise input voltage $E_n/A_n = e_n$. Extreme care must be taken when making the noise measurements. The test circuit should be well shielded to avoid hum pickup and corresponding measurement errors. When large source resistances are used, stray capacitance shunting R_s will cause attenuation of the high-frequency components of the noise voltage, thus reducing the effective noise bandwidth. (b) Equivalent input noise voltage versus source resistance for constant noise bandwidth. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)

at the output of the amplifier, to the theoretical noise power that would be contributed by an ideal resistor equal to the real part of the terminal impedance. It is always greater than unity. The corresponding logarithmic expression is called the noise figure:

$$\text{noise figure} = 10 \log_{10}(\text{noise factor}) \quad (12)$$

The last term of interest is the equivalent amplifier input noise voltage for a constant effective noise bandwidth. It is defined as follows:

$$e_n = \sqrt{(e_{n0})^2 + (i_n r_s)^2} \quad (13)$$

where e = amplifier equivalent RMS input noise voltage

e_{n0} = amplifier equivalent RMS input noise voltage for zero source impedance

i_n = amplifier RMS input noise current

r_s = source resistance

Another source of error in operational amplifiers is the nonideal amplitude and phase response to variations in frequency.

The gain of the circuit is not uniform nor is the phase shift. In critical servosystems this could be a source of instability if not recognized. Some amplifiers also show a change in gain as temperature, time, and supply voltage vary.

The impedance of operational amplifiers seldom approaches the ideal infinite input impedance and zero output impedance. This compromises some of our earlier assumptions in the development of circuit theory. The practical implications are that loading effects must be carefully considered when designing the external input and load circuits. It is important not to approach the saturation voltages of the amplifier or nonlinear response will result.

Common-mode error is associated with operational amplifiers connected in the differential mode. The unit should be responsive only to the difference in voltages applied to the two input terminals (Figure 59). Unfortunately, it tends to amplify each voltage as a separate entity. Since the circuitry at each terminal is minutely different, the net result is a small error called the common-mode error. Let C_n be the ideal differential voltage; then, ideally,

$$C_n = C_B - C_A = \frac{-e}{A} \quad (12)$$

where C_B = voltage applied to the inverting terminal

C_A = voltage applied to the noninverting terminal

e = output voltage

A = gain of the amplifier

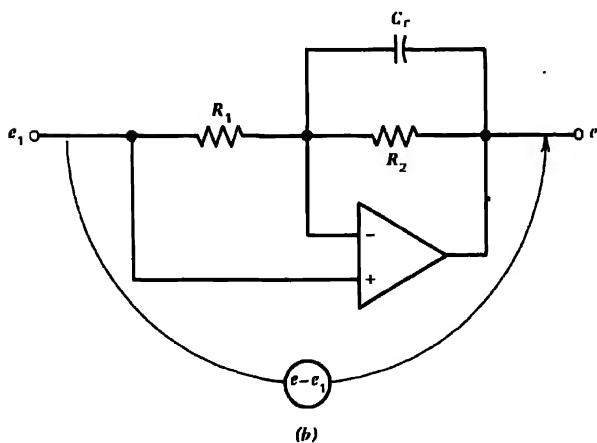
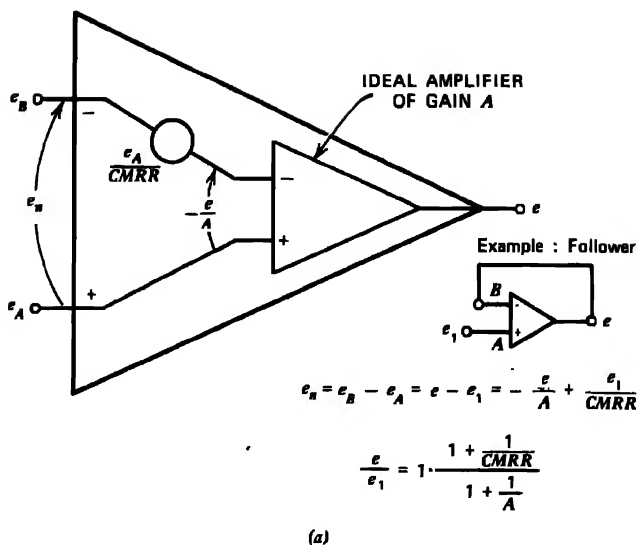


Figure 59. (a) Common mode rejection ratio defined. (b) Measuring common mode rejection ratio. (Courtesy of Philbrick Researchers, Inc.)

By definition, the common-mode voltage is e_A ; some literature also refers to it as e_{CM} :

$$e_{CM} \equiv e_A \quad (13)$$

To define this error quantitatively, the term common-mode rejection ratio (CMRR) is introduced. It is defined in Figure 59b:

$$CMRR \cong \left(\frac{-e_1}{e - e_1} \right) \left(\frac{R_2}{R_1} + 1 \right) \quad (14)$$

It is used in the following equation, which defines the common-mode error:

$$e_n = e_B - e_A = \left(\frac{-e_A}{CMRR} - \frac{e}{A} \right) \quad (15)$$

where e_n = common-mode error.

The CMRR term is usually 20,000 to 200,000. Some literature also uses the expression CMR, which is defined as

$$CMR = 20 \log_{10} |CMRR| \quad (16)$$

It is used as follows:

$$CMR = - \frac{\text{gain of the difference signal}}{\text{gain of the common-mode signal}} \quad (dB) \quad (17)$$

It should be noted that the common-mode error varies with frequency and the magnitude of the signal. When terminal *A* (Figure 59a) is grounded or either signal becomes zero, the error is eliminated.

8.6.4. Electrical Parameters

So far we have discussed the negative side of operational amplifiers. Now we proceed to the positive side—how to specify the desirable electrical characteristics.

OPEN-LOOP GAIN. The open-loop gain is defined as the magnitude of the amplification factor, *A*. It typically lies between 10^3 and 10^6 (Figure 60). Amplifiers are almost never used in this mode because of the extreme sensitivity. An amplifier must operate in its linear range to be useful. The saturation limits are normally slightly less than the supply voltages. If an operational amplifier were used in the open-loop configuration, it would be almost impossible to keep it from saturating with the smallest input voltages.

CLOSED-LOOP GAIN. The closed-loop gain was described and illustrated in Figure 50.

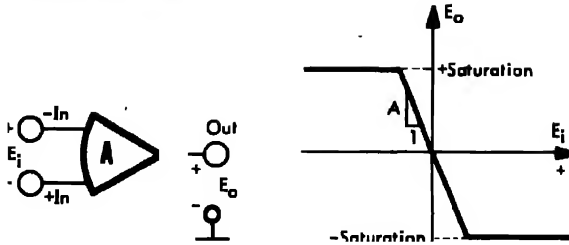


Figure 60. Open-loop transfer curve. Note that the slope of this curve is equal to the gain, A . Since the amplifier operates on finite values of power supply voltage, the input-output curve exhibits a saturation effect at slightly less than power supply voltage. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)

FREQUENCY RESPONSE. The frequency response of an operational amplifier is best represented by a Bode plot (Figure 61). It can easily be constructed when the DC open-loop gain, A_o , and the unity gain bandwidth, f_c , are known. The open-loop gain, $A(j\omega)$, for an amplifier with a 6-dB/octave rolloff can be expressed as

$$A(j\omega) = \frac{A_o}{1 + j(\omega/\omega_o)} \quad (1B)$$

where ω_o = the 3-dB attenuation frequency (radians per second)

$$\omega_o = 2\pi f_o = 2\pi f_c/A_o$$

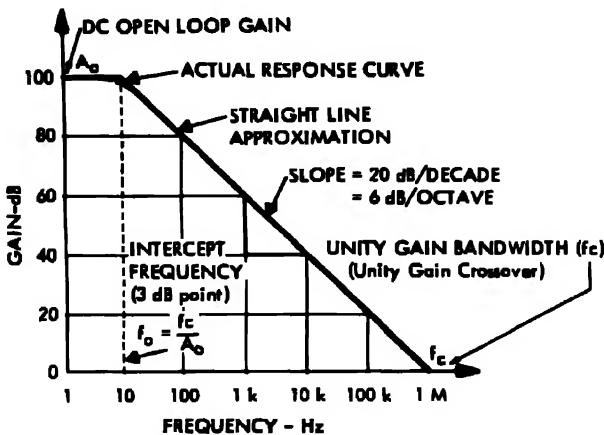


Figure 61. Open-loop gain versus frequency characteristics (Bode plot). (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)

The closed-loop response for a differential circuit can be calculated from the following (Figure 50b):

$$\frac{E_o}{E_i} = \frac{-R_o/R_i}{1 + (1/A\beta)} \quad \beta = \frac{R_i}{R_i + R_o} \quad (19)$$

This is a more rigorous definition than previously presented; it does not assume that A approaches infinity.

The 3-dB frequency for closed-loop gain is called f_{c1} (Figure 62):

$$\text{closed-loop response} = G_c(j\omega) = \frac{(R_o + R_i)/R_i}{1 + j(\omega/\omega_c)[(R_o + R_i)/R_i]} \quad (20)$$

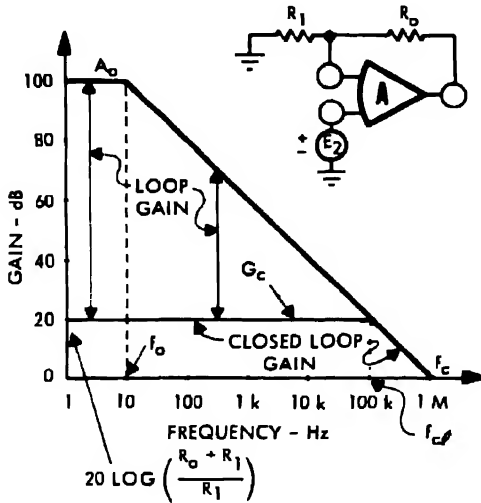


Figure 62. Effect of open-loop gain on closed-loop response. (Courtesy of Burr-Brown Research Corporation, Inc. From Reference 27.)

A plot of $G(j\omega)$ shows that at low frequencies, where loop gain is high, the closed gain is determined by the feedback network. At high frequencies, where loop gain is small, the open loop is the determining factor and the closed-loop plot merges with the open-loop plot. Furthermore, it can be shown that the gain-bandwidth product is a constant at any closed gain:

$$(f_{c1})(G_c) = \text{constant} = (f_o)(A_o) = f_c \quad (21)$$

Note that gain is expressed in decibels.

SLEWING. Slewing is the maximum rate at which the amplifier can be driven without producing distortion or nonlinear effects. It is expressed in change of voltage per unit time.

SETTLING TIME. This is defined as the period required for the amplifier to reach 63% of the output voltage. It is measured in microseconds or nanoseconds.

8.6.5. Selection

The array of operational amplifiers today is tremendous. They can be categorized by physical construction, type of input connections, and electrical parameters. As applications for these devices increase, manufacturers are quick to capitalize on the commercial opportunities to build a special amplifier. The commonly stocked amplifiers listed by electrical function are as follows:

1. Ultralow drift amplifier has maximum drift voltages of $\pm 1 \mu\text{V}/^\circ\text{C}$.
2. Low-noise amplifier maintains the noise as low as $2 \mu\text{V}$ (DC to 1 kHz).
3. Universal amplifiers feature high common-mode rejection and insensitivity to power-supply fluctuations.
4. FET (field effect transistor) input amplifier has characteristically high input impedances, low-noise characteristics, and high slewing rates.
5. Chopper-stabilized amplifiers feature minimum drift rates and low-noise characteristics.
6. Wide-band amplifiers have bandwidths from DC to 10 MHz. They frequently utilize FET inputs and chopper stabilization.
7. High-voltage and high-current amplifiers offer output voltages up to 115 V and currents up to $\frac{1}{2}$ A.
8. Power boosters are unity-gain amplifiers that are used in series with the output of other amplifiers. They enable the combination to supply more than the normal power. The input impedance of the device is high and the output impedance very low; this results in low driving currents and relative freedom from loading effects.

The second generally accepted way of categorizing operational amplifiers is by physical construction. There are discrete-component, monolithic, and hybrid amplifiers.

1. Discrete-component units are built around printed circuit boards, solid-state components, and all the necessary auxiliary bypass capacitors and rolloff networks. They are normally encapsulated to provide environmental and handling protection. They can be used economically in small quantities. Some varieties are also available with vacuum tubes for special purposes. It is relatively simple to change components when optimizing a design.

2. Monolithic amplifiers have no discrete components; instead all components are combined in a series of layers. This process results in high tooling costs when nonstandard units are required. However, the unit price is low. Since so many different types of monolithic units are now catalog items, a properly engineered design is inexpensive. The package is extremely small, typically the size of a TO-5 package transistor or an inline package. They may require external circuitry for stabilization networks. They are presently considered to be the most reliable type of amplifier, since the interconnections between components have been eliminated.

3. Hybrid amplifiers are a cross between discrete and monolithic units. They use both types of components and are available in a package size not much larger than monolithic units. Their reliability is considerably better than that of discrete component units, but not so good as that of monolithic units.

8.6.6. Glossary

A glossary of operational amplifier terms appears in Appendix E.

REFERENCES

1. C. Dawes, *Electrical Engineering*, Vol. 1, 3rd ed., McGraw-Hill Book Company, New York, 1937.
2. F. W. Sears, *Principles of Physics*, Vol. 2, Addison-Wesley Publishing Company, Cambridge, Mass., 1946.
3. H. C. Roters, *Electromagnetic Devices*, John Wiley & Sons, Inc., New York, 1966.
4. Mechanical Drives Issue, *Machine Design Magazine*, December 18, 1969, Penton Publishing Company, Cleveland, Ohio.
5. V. L. Maleev, *Machine Design*, International Book Company, Scranton, Pa, 1946.
6. *System Designers Handbook*, January 1965, Benwill Publishing Company, Brookline, Mass.
7. E. G. Jackson, *Lubrication in Space Vehicles*, National Research Corporation, 1962.
8. J. D. Alber and R. H. Ingrau, Environmental Problems in External Space Equipment, *IEEE Spectrum*, July 1965.
9. W. A. Salmon and C. M. Apt, *A Lubrication System for Space Vehicles*, Society of Automotive Engineers, January 1963.
10. R. L. Johnson and W. J. Anderson, *Summary of Lubrication Problems in Vacuum Environment*, Lewis Research Center, NASA, Cleveland, Ohio, 1962.
11. R. L. Johnson and D. H. Buckdry, *Lubricants and Mechanical Components of Lubrication Systems for Space Environment*, Lewis Research Center, ASLE, Annual Meeting, Chicago, May 1964.
12. F. J. Williams, *Design Engineers View of Solid Dry Film Lubricant Coatings*, ASME Symposium, June 1965.

13. S. F. Murray, P. Lewis, and A. J. Babecki, *Lubricant Life Tests on Ball Bearings for Space Applications*, ASLE, May 1966.
14. D. M. Winalow, D. Horowitz, and D. V. McIntyre, Adhesion and Cohesion of Metals in Space Environment, *Journal of ASLE*, June 1966.
15. S. Hanau and S. Abromowitz, *New Concepts in Bearing Designs and Applications*, International Techtonics, Inc., Compton, Calif., 1961.
16. C. E. Norton and D. J. Baker, *Effect of New Materials and Processing Developments on Ball Bearing Fatigue Life*, Society of Automotive Engineers, January 1964.
17. H. C. Rippel and J. S. Tawresy, *An Analytical Study of Bearing Noise and Vibration Generation and Transmission*, Franklin Institute Report I-A 2321-1, 1958.
18. C. H. Hannan, *Progress Report on Barrier Film Work*, MPB Corporation, 1968.
19. F. M. Yana and N. R. Gardner, Beryllium—The Metal and its Design Potential, *Electrical Manufacturing*, November 1959.
20. *A FBMA Standards for Instrument Ball Bearings*, Section 12, The Anti-Friction Bearing Manufacturers Association, New York, 1955.
21. Power Supply Issue, *Electronic Design*, April 19, 1966, Hayden Book Company, Inc., New York.
22. *SCR Manual*, General Electric Company, Auburn, N.Y., 1964.
23. P. Birman, *Power Supply Handbook*, Kepco, Inc., Flushing, N.Y., 1965.
24. P. Birman, *Operational Power Supply Technology*, Kepco, Inc., Flushing, N.Y., 1969.
25. K. Olson, Power Supplies, *Electromechanical Design*, July 1966, Benwill Publishing Company, Brookline, Mass.
26. *Operational Amplifiers*, Radiation, Inc., Melbourne, Fla., 1962.
27. *Handbook of Operational Amplifiers*, Burr-Brown Research Corporation, Inc., Tucson, Ariz., 1969.
28. *Handbook of Operational Amplifier Active RC Networks*, Burr-Brown Research Corporation, Inc., Tucson, Ariz., 1969.
29. *Philbrick/Nexes Operational Amplifier Catalog*, Dedham Mass., 1965.
30. *Applications Manual for Computing Amplifiers*, Philbrick Researchers, Inc., Dedham, Mass., 1966.
31. R. F. Shea, *Amplifier Handbook*, McGraw-Hill Book Company, New York, 1966.
32. S. A. Walston and J. R. Miller, Ed., *Transistor Circuit Design*, McGraw-Hill Book Company, New York, 1963.

Chapter IX System Integration

This chapter examines system integration in its broadest aspects—the collective elements that frame a successful project. Some of the material is purely technical and a small segment is subjective; all of it is based on actual engineering experience.

9.1. DEFINING GOALS

Defining the objectives of a system is often more than 50% of the task at hand. After the contract has been awarded, the salesman and top brass directed toward new horizons, and the air cleared of often recited past company glory, it becomes the job of the systems engineer to establish concrete technical objectives. Very often the voluminous contracts that precede awards are more explicit about methods of payment, termination procedures, progress reports, and F.O.B. points than the technical details of the tasks involved. A practical first approach is an analysis of the difference between what the customer asked for and what is really required to perform the task. Since the cognizant engineer is recognized as the expert in the field of the new device, it is his responsibility to eliminate any mutual or unilateral misunderstandings before the project starts. After these initial steps are carefully taken, it is possible to set design objectives and establish the parameters and tolerances required by the system and its components.

9.2. SCREENING COMPONENT DATA

After system analysis to determine the overall component parameters, a basic choice must be made between standard commercial components and "MIL-Spec." units. In military programs "MIL-Spec." parts are prescribed.

The military systems and all components are governed by rigid specifications such as MIL-E-5272, MIL-E-5400, and numerous supporting documents. The importance of these specifications is that they completely define component environmental performance and a great deal of the design details; little is left to chance. A component certified to the applicable military specification is usually, but not always, of the highest quality. In return for this type of insurance the system designer pays a substantially higher price, longer delivery times are required, and the number of qualified manufacturers is limited. In some instances the only difference between military and commercial components are a label and the price. One of the objectives of this book is to provide an outline for evaluating each component so that the engineer can properly determine its dollar value.

Most nonaerospace systems are equipped with commercial hardware. The chief problem in selecting and comparing competitive components is to determine the basis of the tolerances listed in the manufacturers' catalog. For example, if a potentiometer is listed as having an accuracy of 1%, the following questions arise:

What is the basis of the number?

Was independent or terminal linearity used?

Was the test based on room temperature data or an extended temperature range?

Was the evaluation performed under conditions of shock and vibration?

How would extended usage affect the tolerance?

What effect would humidity and normal dust have on the component?

Before the units can be properly compared, all of these points must be clarified. Another equally important question is "What tolerances and environmental conditions are significant for the project at hand?"

When the army purchases a solenoid rated at 18 to 32 V and a given lifting capacity at a prescribed stroke, it inspects the solenoid for worst-case conditions. The unit is cycled at 32 V for a given number of hours and then tested for lifting power at 18 V. The thermal load is thus maximized and the pulling power minimized. The test document also prescribes the test fixture design to standardize the thermal conditions. This is an extremely severe acceptance test and no catalog item could pass it; yet this is the system requirement. If the usage requires this type of test, catalog information is useful only as a start toward further development of this kind of solenoid.

Most American component manufacturers publish an exceptionally good technical digest of their products. However, it is not possible to provide data for all conditions. This must be constantly kept in mind when reviewing technical catalogs.

9.3. ENVIRONMENTAL FACTORS

This section reviews collectively the effects of environmental factors on component performance. This should assist the systems engineer in making an "apple-to-apple" comparison of components.

9.3.1. Hot and Cold Environments

The principal effects of temperature variations are the following:

1. Change in resistance of wound components, and electrical parameters in semiconductors.
2. Changes in linear dimensions.
3. Changes in mechanical properties, such as strength.

As temperatures vary, the resistance of the coil in wound components varies and the associated heating, response time, and power output also change. Consequently, the most severe test for these types of units is at high temperatures; low temperatures will never produce thermal problems. Although the change in power levels under stabilized conditions is a recognized design problem, a more subtle effect is the change in characteristics as the unit is heating or cooling. This can cause servo instability since the response time is also affected. It can also occur as a result of varying convection patterns.

Changes in linear dimensions chiefly affect mechanical fits and clearances. Most electromechanical devices are designed so that the majority of mating metal parts are made of similar materials; in this way the parts expand and contract at the same rates when subjected to thermal cycling. However, it is not always possible to achieve this design ideal. In most instruments the end play, or axial clearance, can be expected to vary with temperature. Bearings as an entity are not affected by changes in internal clearances but the parts they are pressed into, or the shafts they support, may be plagued by the differential expansion problem. The principal thermal effect in bearings is the change in torque due to changes in viscosity in the lubricant. The critical test for bearing torque is therefore at low temperatures. Gear trains are also a source of trouble. When gears of different materials are mated, the clearances between gears may change, resulting in increased torque readings.

As temperature variations exceed the conventional limits of -55 to $+200^{\circ}\text{F}$, the physical characteristics of the materials must be considered; this is particularly true when plastic components are used. Parts that are pressed into other members will be subjected to additional stresses. There will be changes in ultimate strength, elastic and endurance limits, variations

in ductility, impact strength, and modulus of ductility. Mechanical hysteresis is also increased at elevated temperatures. One of the most insidious high-temperature effects is creep. It is particularly troublesome in components that use bellows or diaphragms. Most of these use solder for sealing, which can produce undesirable effects at temperatures as low as 70°C.

High ambient temperatures are not the sole cause of thermal failure; in most cases poor heat-sinking techniques are equally responsible. Any electrical component that continuously generates heat that is not conducted or convected to a heat sink will overheat. Whether the component is a 1 W carbon resistor mounted on insulating plastic or a heavy-duty motor mounted on a plastic base, the problem is basically the same. The heat-generating element must be provided with a thermal path to the heat sink. The mass of the heat sink will appreciably influence the temperature of the component and, consequently, the performance. For example, consider the case where a size 20 (2.00 in. in diameter) servomotor is tested in an oven maintained at 100°C. If the motor is suspended by a wire from the top of the oven, thermal conduction of the heat generated is impossible and we can anticipate overheating effects. The other extreme would be to mount the motor on a 1-ft² aluminum block where the heat sinking is ideal; this would not be realistic either. The only fair test is one simulating the intended use, and this is not usually found in catalogs. Temperature ratings found in manufacturers' literature must be evaluated on the basis of the heat sink they used. Convection effects must similarly be considered.

Another serious problem associated with high temperatures is outgassing. Outgassing is the generation of gases from volatile components of a substance. For example, some of the varnishes and potting materials used on wound components produce relatively large volumes of gases during the break-in period. All types of paints, lacquers, and low-temperature plastics produce gaseous products under very moderate heat. Low-temperature insulation, residual soldering flux, and some oils are also problems. This effect is greatly aggravated at low pressures, such as those found in aircraft and space vehicles. The chief problem associated with outgassing is that the gases encounter cooler sections of the system, reform into liquids or solids, and deposit on critical elements such as shafts, pivots, and bearings. The net effect is an increase in operating torque. A second effect is the deposition of gases on electrical components where they may decrease insulation resistance between electrical components, or possibly cause shorts. These deposits have also been known to seal small pressure ports or venturis used in pneumatic or hydraulic equipment. Otherwise competitive components frequently have very different outgassing characteristics. This problem is particularly important in optical systems where any deposition on lenses or mirrors has a catastrophic effect.

9.3.2. Shock Loading

The fundamental nature of shock loading is to cause a distortion, or failure of components, by exceeding the elastic limit or the ultimate strength of some instrument element. The biggest problem in designing systems subject to shock loading is determining the probable shock wave pattern or shock signature. Five basic signatures are found in current literature and catalogs:

1. *Impulse.* The shock increases from zero to a finite amplitude in a negligible period of time; the acceleration is maintained for a given period of time and then decreases to zero in a negligible amount of time. It is sometimes called a square-wave shock. A popular modification of the square wave is the trapezoidal pulse. It differs from the square wave only in terms of the rise and decay times of the pulse, which are finite and prescribed. Trapezoidal shock waveforms are relatively popular because they can easily be produced on most shock machines.

2. *Step.* The shock wave increases from zero to a finite acceleration in a negligible amount of time and is maintained at that level for a relatively long period of time. The only practical way of producing this type of waveform is on a double centrifuge. This shock wave is rarely used because of the scarcity of special equipment required to synthesize it.

3. *Half-sine.* This shock signature is the most popular in current technology. The pattern follows a sine wave from zero to π radians and then remains at zero for π radians. It can be produced by most shock machines in use today. The prime differences between half-sine shock pulses is the duration of the pulses; 11 msec is the most commonly used duration but 2, 8, 18, and 22 msec are also used. The longer the pulse duration, the greater the energy in the pulse.

4. *Decaying sinusoid.* Decaying sinusoid shock pulses represent a more realistic approach to shock synthesis than single-shock pulses. This method attempts to deal with the fact that most objects are second-order systems with a specific damping ratio and tend to continue vibration in a damped sinusoidal mode after the application of a single pulse. The degree of damping must be carefully estimated to make this approximation reasonably correct. The test equipment required to produce this shock wave is an electrodynamic vibrator programmed by some type of tape input. It is presently used only in very sophisticated systems.

5. *Complex.* Complex waveforms go one step further than damped sinusoidal techniques by recognizing that most shock waves found outside of the test laboratory are a series of Fourier-type waveforms that are random and complex. The only way of producing this type of shock pulse is to use an electrodynamic vibrator programmed with a tape that has been derived

from actual field experiments. It is closely allied with random vibration work and is commonly used in aerospace system testing. The waveform is usually identified by at least one peak value of acceleration for a given period, in addition to standard random vibration terminology.

Few manufacturers of electromechanical components provide test data more extensive than under a half-sine wave or trapezoidal signature. The commercial approach is usually modest field testing; the military project will normally prescribe a specific shock pulse. The brunt of shock loading is applied to bearings, supporting shafts, and torsion bars. When selecting components for a shock environment it is usually a good idea to select a unit with somewhat oversized supports. This is more economical than an extensive test program.

9.3.3. Vibration

The selection of electromechanical components to be used in a vibration environment is usually less concerned with the survival of the unit than with how the forcing frequencies will modulate the output of the system. It is reasonable to assume that most catalog hardware is capable of resisting linear vibration to the extent listed in the specification. Survival under random vibration should be carefully checked.

9.3.4. Acceleration

The application of steady-state acceleration is rarely a source of component failure; it does cause parasitic modulation. Any part in a system that is not balanced will be affected by acceleration. All gears, drums, and shafts that are not statically balanced will be caused to rotate under constant acceleration. The action is similar to that on a simple pendulum where acceleration acting on a mass, at a fixed radius from the center of rotation, causes it to deflect. Most good rotary devices such as motors, gyros, tachometers, and navigational equipment are statically and dynamically balanced. The auxiliary devices coupled to them are often neglected.

9.3.5. Altitude

High altitude or low pressure tests are designed to check the integrity of gaskets and seals and the degree of outgassing of components. A sealed system can be severely damaged mechanically by a sudden leak in a gasket

at high altitudes. A secondary effect is to considerably increase the degree of outgassing. The viscosity of fluids may also be affected by accelerating the vaporization of lighter elements in the mixture. If pressure seals are an important system consideration, some form of leak test should be specified.

9.3.6. Corrosion

Military specifications have established a precedent in requiring many systems and components to withstand humidity and salt spray tests. Although essential for military equipment, salt spray tests are difficult to justify on commercial equipment not intended for marine environments. The cardinal point in all such tests is how well the plating on metal parts withstands this environment. Plating that cracks, peels, or is porous will rapidly disintegrate. An extra 10 cents spent on good plating will result in more than a dollar in performance. The spread between good and marginal plating is normally a factor of 2 or 3, but this should be the last place to try to economize. Commercial practice has been to use humidity tests to spot-check the quality of plating, since the equipment is inexpensive and the results reliable.

9.3.7. Contamination

All components are susceptible to damage by contamination to some degree. The Achilles' heel of mechanical parts is the bearing; electronic components are vulnerable at the solder joints. External contaminants may be solids, liquids, or gases. Internal contaminants may be parts that flake or products of outgassing. The standard approaches to keeping contaminants out of components are dust seals, gaskets, and solder (hermetic) seals. The majority of high-performance commercial equipments use gasket seals since they are effective and inexpensive. Very precise gyros, accelerometers, and military hardware use hermetic seals with an enclosed atmosphere of dry nitrogen and helium. Some components, such as miniature relays, are used in so many military applications that hermetic sealing has become standard. Very often the most dangerous contaminants are built into the component during assembly. These may be dirt, hair, shreds of material, metal chips, loose solder particles, flux and plastics that outgas, unplated parts that oxidize and subsequently flake, and, sometimes, humid air. White room techniques have substantially scaled down many of these problems, but, unfortunately, they are still the most common causes of component failures. The only defense against contamination failures is a very vigilant quality-control department.

9.4. ELECTRICAL CONSIDERATIONS

When designing an electromechanical system, there are many considerations other than circuitry and hardware. For example:

1. Compatibility with other systems and hardware.
2. Replacement parts.
3. Servicing.

Vacuum tubes are rarely considered for new designs today, except where high power at high frequencies is required. Nevertheless, many systems, components, and appliances still use tubes. Microwave units, television sets, and machinery controls are typical examples. The availability of power supplies, standard components, and trained personnel cannot be ruled out without considerable thought. In many of these applications a small package size is of no particular interest. For example, an amplifier reduced from a 4 × 8 in. chassis to the size of a thumbnail has little significance in a power plant. The fact that the available service personnel are still oriented toward vacuum tubes may be more significant. The array of special-purpose and multifunction tubes is still a significant factor.

The bulk of modern circuit design is done with discrete solid-state components. They combine versatility, moderate power capacity, availability, and small envelopes. Like vacuum tubes, transistors mounted in sockets are easy to change when one goes bad or the circuit is changed. In contrast to integrated circuits, it is a simple matter to switch and optimize each segment of an amplifier. The development of circuit boards specifically for discrete components with auxiliary heat sinking members has increased the popularity of this approach. Standardized circuitry, too, is abundantly available, mounted on circuit boards. Almost any conceivable amplifier, oscillator, buffer, and hundreds of other circuits are available in optimized form. If the stock circuits match the system requirements, the savings are impressive; very often, when the match is not exact, the system is slightly modified to effect the cost saving. The majority of the packaged circuits now available have a long history in service, and much reliability data is available. This is never true of new circuitry no matter how brilliantly conceived. The biggest advantage of discrete components over integrated circuits is power rating. Another form of packaged discrete components is the module. This consists of a number of components wired together and mounted on a solid member such as a cone or bar. The structure provides rigidity, heat sinking, and a simple way of packaging an entire circuit. The module is mounted in a socket and can easily be replaced. This technique is older than mounting circuits on a flat circuit board but it is still in wide use, particularly where power levels are high and good heat sinking is important. It occupies more space than a series of flat boards but convection cooling is more practical.

Integrated circuits provide the ultimate in circuit volumetric efficiency. They are intended primarily for logic work rather than power. Compared to discrete components, integrated circuits may be classified as the "brains" while transistors provide the brawn. The number of complete circuits available in this form far surpasses the number of those available in the discrete form. In logic circuitry design they are considered to be the basic building blocks in the same manner as resistors, capacitors, inductors, and transistors are used in other circuits. Special logic circuitry that does not employ "catalog-type" integrated circuits is extremely rare today. Although integrated circuits have been directed toward the computer industry, many special circuits are available wherever the market requires them. For example, the computer industry principally utilizes NAND, NOR, flip-flops, and driver circuits where servoamplifiers may be considered a separate entity. Circuits based on operational amplifiers are particularly popular; many manufacturers make exactly the same circuit with resultant savings for the circuit designer. One of the best features of integrated circuits is that they include all the circuit safety devices that designers frequently omit in discrete circuits.

Tubes, transistors, and integrated circuits, when compared on the basis of compatibility with other equipment, have specific virtues. Vacuum tubes are most capable of withstanding electrical abuse, such as abnormally high voltage and transients; solid-state components must be protected against these problems. When the abuse is mechanical, solid-state components excel; they easily withstand shock and vibration that would cause tubes to break. Besides high-frequency equipment, tube circuits are most adaptable to motor controls where high transients are normal. If the new system is to be married to an existing system, keep in mind that tube and solid-state power supplies are considerably different.

Replacement parts for tube, discrete components, and integrated circuits are not likely to be a problem if the designer selects standard components. The current selection of modern vacuum tubes is not likely to disappear from the market for many years. Transistors are constantly being improved; when an older line disappears from current stocks, an acceptable replacement is normally available. The only dangerous situation is when a special transistor is ordered from a small company. The situation with integrated circuits is about the same. If the circuit designer uses "popular" items, particularly those made by more than one manufacturer, the probability of future replacement is excellent. Most large manufacturers, as well as the military, periodically print a list of preferred circuits, components, and modules; it is a good idea to check these lists at the inception of any system design.

Servicing of electromechanical systems is a problem as old as the industry. Once the system must be field-repaired, its chances of ultimate survival

diminish sharply. There are exceptions, of course. Some large firms establish repair depots at key locations where specially trained personnel is available. Another approach often used is to supply repair manuals containing a list of trouble symptoms and how to eliminate them. Replaceable circuit board construction is particularly valuable here because the system can be analyzed in a manner similar to "television repairing." When the problem involves troubleshooting individual components, it is likely that a man familiar with vacuum tube circuits will be available. Technicians with transistor experience will be a close second, and those with integrated circuit familiarity a poor third. Ten years from now the ratios will be decidedly different.

9.5. ECONOMIC CONSIDERATIONS

A successful system designer is one who produces a device that is technically acceptable at a cost that results in a profit. A good proposal is a prerequisite for any enterprise. A complete discussion of the preparation of a proposal is beyond the scope of this chapter but a few of the common errors are worth discussing.

When estimating the cost of a component, all the cost-contributing tolerances and environmental requirements must be considered. Suppose that we use a potentiometer as an example. The problem is not so much specifying a 1% linearity unit when 0.1% is actually required, but noting that the unit may be required to function at 100°C without degradation of performance or in a 0.1 G^2 /Hz vibration environment without contact bounce. All of these requirements have a substantial impact on price. Conversely, never use a precision pot when a standard unit is adequate, since this gives the competition an advantage. An intelligently designed circuit uses only standard components; avoiding special tool costs is a prime rule for commercial work. In general, there are no bargains in electro-mechanical devices; you get only what you pay for. However, there are some exceptions. By checking with local distributors, the purchasing department frequently can find surplus standard units produced for a large order that was not fully utilized, such as a four-input NAND circuit manufactured for a computer program or a precision resolver intended for a canceled gunfire director program. Such items are in no way inferior to stock units.

9.6. SYSTEM ERROR ANALYSIS

One of the most important tasks performed by the systems engineer is to predict the overall error of a complex electromechanical device. There are

numerous mathematical approaches to this problem; the choice is somewhat subjective. The method listed below has proved to be very useful for the majority of electromechanical problems.

The general approach is to analyze the system as a group of inputs that are added at one central summing point. The mathematical background is as follows (Figure 1).

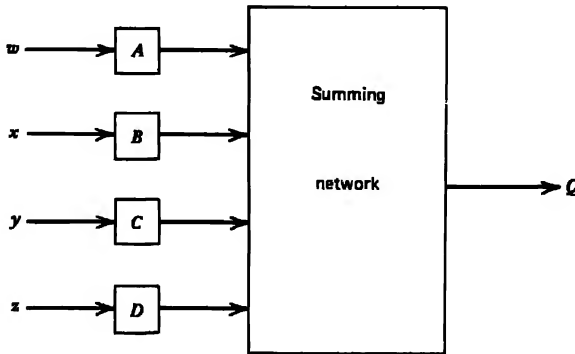


Figure 1. General approach to error analysis.

Given a summing network, with inputs w , x , y , and z , from transducers with transfer functions A , B , C , and D , the output Q is

$$Q = Aw + Bx + Cy + Dz \quad (1)$$

The total differential in terms of its partials is

$$dQ = \frac{\partial Q}{\partial A} dA + \frac{\partial Q}{\partial B} dB + \frac{\partial Q}{\partial C} dC + \frac{\partial Q}{\partial D} dD \\ + \frac{\partial Q}{\partial w} dw + \frac{\partial Q}{\partial x} dx + \frac{\partial Q}{\partial y} dy + \frac{\partial Q}{\partial z} dz \quad (2)$$

Since we are determining the error in Q , Q is a function of the variations in the inputs and the equation may be rewritten as follows:

$$\Delta Q = \frac{\partial Q}{\partial A} \Delta A + \frac{\partial Q}{\partial B} \Delta B + \frac{\partial Q}{\partial C} \Delta C + \frac{\partial Q}{\partial D} \Delta D \\ + \frac{\partial Q}{\partial w} \Delta w + \frac{\partial Q}{\partial x} \Delta x + \frac{\partial Q}{\partial y} \Delta y + \frac{\partial Q}{\partial z} \Delta z \quad (3)$$

The equation that governs Figure 1 is equation 1:

$$Q = Aw + Bx + Cy + Dz$$

The partial derivatives are

$$\frac{\partial Q}{\partial A} = w \qquad \frac{\partial Q}{\partial w} = A$$

$$\frac{\partial Q}{\partial B} = x \qquad \frac{\partial Q}{\partial x} = B$$

$$\frac{\partial Q}{\partial C} = y \qquad \frac{\partial Q}{\partial y} = C$$

$$\frac{\partial Q}{\partial D} = z \qquad \frac{\partial Q}{\partial z} = D$$

They are also called influence coefficients. The procedure for using this approach is as follows:

1. Assign numerical values to each of the partial derivatives based on physical conditions. Since A , B , C , and D are the gains of transducers, they may be expressed as a given number of volts per volt, volts per foot, volts per foot per second, volts per foot per square second, volts per degree Fahrenheit, volts per psi, or any other common transducer terminology. All of these illustrations pertain to electrical servosystems; mechanical summing networks might appropriately be foot-pounds per foot per second, foot-pounds per degree Fahrenheit, or any other of the constants typical of pneumatic, hydraulic, or mechanical servosystems.

2. Determine the reasonable range of values for each of the independent variables, w , x , y , and z . For example:

$$5 \leq w \leq 25 \text{ ft}$$

$$20 \leq x \leq 40 \text{ ft/sec}$$

$$1 \leq y \leq 2 \text{ ft/sec}^2$$

$$0 \leq z \leq 200^\circ\text{F}$$

3. Compute the average value of each input and use that as its most probable value. This is based on the assumption that the normal distribution curve for each is a regular bell-shaped curve:

$$w = 15 \pm 10$$

$$x = 30 \pm 10$$

$$y = 1.5 \pm 0.5$$

$$z = 100 \pm 50$$

Let us assume that the error associated with each variable is 1% of the nominal value.

4. Estimate the gain and the errors inherent in the transducers by checking the manufacturers' catalogs. They might be as follows:

$$A = 5 \pm 1\%$$

$$B = 20 \pm 0.1\%$$

$$C = 0.5 \pm 0.01\%$$

$$D = 10 \pm 0.5\%$$

5. Tabulate the absolute value of the various errors:

$$\Delta A = 0.01 \times 5 = 0.05$$

$$\Delta B = 0.001 \times 20 = 0.02$$

$$\Delta C = 0.0001 \times 0.5 = 0.00005$$

$$\Delta D = 0.005 \times 10 = 0.05$$

$$\Delta w = 0.001 \times 15 = 0.015$$

$$\Delta x = 0.001 \times 30 = 0.030$$

$$\Delta y = 0.001 \times 1.5 = 0.0015$$

$$\Delta z = 0.001 \times 100 = 0.1$$

6. Take the root-sum-square of the sum of the products of each error and its applicable influence coefficient (Table 1).

7. Compute the nominal value of Q from equation 1:

$$Q = 15 \times 5 + 30 \times 20 + 1.5 \times 0.5 + 100 \times 10 = 1675.75$$

8. Compute the average system error:

$$\frac{\Delta Q}{Q} = \frac{5.22}{1676} = 0.31\%$$

For more complex configurations see Figure 2.

In addition to predicting average system error, this technique shows which transducer and input have the largest effect on overall performance.

The following problem is presented to demonstrate the effect of environmental factors on component selection. A summing network was designed to compute electrically the product of two variables plus one additional input (Figure 3):

$$Q \propto x + yz \quad (4)$$

Table 1. Computation of RMS Error

Parameter	Absolute Magnitude	Influence Coefficient	Error
ΔA	0.05	$\frac{\partial Q}{\partial A} = 15$	0.75
ΔB	0.02	$\frac{\partial Q}{\partial B} = 30$	0.60
ΔC	0.00005	$\frac{\partial Q}{\partial C} = 1.5$	0.000075
ΔD	0.05	$\frac{\partial Q}{\partial D} = 100$	5.00
Δw	0.015	$\frac{\partial Q}{\partial w} = 5$	0.075
Δx	0.030	$\frac{\partial Q}{\partial x} = 20$	0.600
Δy	0.0015	$\frac{\partial Q}{\partial y} = 0.5$	0.00075
Δz	0.1	$\frac{\partial Q}{\partial z} = 10$	1.00
ΔQ (root-sum-square) = 5.22			

The A , C , and D may be considered to be transducers; multiplier E is a unity-gain servomultiplier, and B is a scale factor potentiometer. The complete equation is

$$Q = xAB + (yC)(zD)E$$

or

$$Q = xAB + yzCDE \quad (5)$$

The range of independent variables is as follows:

$$\begin{array}{ll} 0 < x < 20 & x = 10 \pm 10 \\ 0 < y < 2 & y = 1 \pm 1 \\ 0 < z < 10 & z = 5 \pm 5 \end{array}$$

The accuracy of all inputs is within $\pm 1\%$ under all environments. The system components under consideration are shown in Table 2.

Transformation	Original diagram	Equivalent diagram	Equation
(a) Combining blocks in cascade			$M_2 = M_1 G_1 G_2$
(b) Eliminating a feedback loop			$M_2 = M_1 \frac{G}{1 \pm HG}$
(c) Eliminating a forward loop			$M_2 = M_1 (G_1 \pm G_2)$
(d) Removing a block from a feedback loop			$M_2 = M_1 \left(\frac{G}{1 \pm HG} \right)$
(e) Removing a block from a forward path			$M_2 = M_1 (G_1 \pm G_2)$
(f) Moving a pick-off point ahead of a block			$M_2 = M_1 G$
(g) Moving a pick-off point beyond a block			$M_2 = M_1 G$
(h) Moving a summing point ahead of a block			$M_2 = M_1 G \pm M_3$
(i) Moving a summing point beyond a block			$M_2 = [M_1 \pm M_3] G$
(j) Moving a pick-off point ahead of a summing point			$M_2 = M_1 \pm M_3$
(k) Moving a pick-off point beyond a summing point			$M_2 = M_1 \pm M_3$
(l) Rearranging summing points			$M_2 = M_1 \pm M_3 \pm M_4$

Figure 2. Block diagram transformation theorems. (Courtesy of McGraw-Hill Book Company. From Reference 2.)

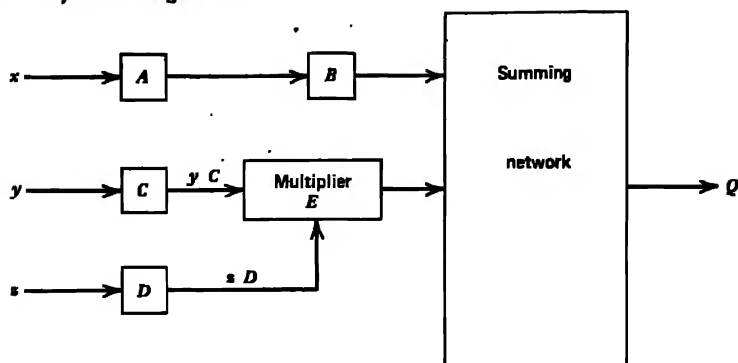


Figure 3. Effect of environmental factors on a summing network.

The object of this problem is to determine the room temperature errors, the effect of environmental factors, on this number and to decide if premium units are justifiable. The influence coefficients are as follows (Table 3):

$$\frac{\partial Q}{\partial x} = AB$$

$$\frac{\partial Q}{\partial B} = xA$$

$$\frac{\partial Q}{\partial y} = zCDE$$

$$\frac{\partial Q}{\partial C} = zyDE$$

Table 2. List of Component Tolerances—Room Temperature

Component	Gain	Tolerance (%)	Hot and Cold (%)	Hot, Cold, Shock, and Vibration (%)
Standard transducer A	2	± 1	± 2	± 3
Premium transducer A	2	0.5	0.6	0.7
Standard transducer C	4	5	6	6
Premium transducer C	4	1	2	2
Standard transducer D	2	5	7	8
Premium transducer D	2	1	2	2
Standard potentiometer B	0.5	1	2	2
Premium potentiometer B	0.5	0.1	0.2	0.2
Standard multiplier E	1	0.1	0.1	0.1
Premium multiplier E	1	0.01	0.05	0.07

Table 3. Standard Components—Room Temperature

Parameter	Error Magnitude	Influence Coefficient	Resulting Error
ΔA	± 0.02	5	± 0.10
ΔB	0.005	20	0.10
ΔC	0.20	10	2.00
ΔD	0.10	20	2.00
ΔE	0.001	40	0.04
Δx	0.10	1	0.10
Δy	0.01	40	0.40
Δz	0.05	8	0.40
ΔQ (root-sum-square) =			± 2.86

$$\frac{\partial Q}{\partial z} = yCDE$$

$$\frac{\partial Q}{\partial D} = zyCE$$

$$\frac{\partial Q}{\partial A} = xB$$

$$\frac{\partial Q}{\partial E} = zyCD$$

The nominal value of Q from equation 2 = 50:

$$\text{system error} = \pm \frac{2.86}{50} = \pm 5.72\%$$

The analysis above shows that the greatest error is contributed by transducers C and D . If money is available, these are the components that should be replaced with premium units. At this point it is equally important to determine if the system requires errors smaller than $\pm 5.52\%$. Let us assume that it is desirable to improve the system to $\pm 3\%$ accuracy under all environmental conditions and that money is available for two premium components (Table 4).

Table 4. Two Premium Components—All Environments

Parameter	Error Magnitude	Influence Coefficient	Resulting Error
ΔA	± 0.06	5	± 0.30
ΔB	0.01	20	0.20
ΔC (premium)	0.08	10	0.80
ΔD (premium)	0.04	20	0.80
ΔE	0.001	40	0.04
Δx	0.10	1	0.10
Δy	0.01	40	0.40
Δz	0.05	8	0.40
ΔQ (root-sum-square) =			± 1.32

Then

$$\text{system error} = \frac{\pm 1.32}{50} = \pm 2.64\%$$

If it is desirable to improve the system so that its error is comfortably below $\pm 2\%$, premium units for transducers *A* and *B* should be selected. This would constitute an optimum price-performance level because any additional improvement would require relatively large expenditures. It would involve component *E* as well as some method of controlling the accuracy of inputs *y* and *z*. The input parameters are normally a function of the customer's equipment, which could lead to nonengineering "complications."

Perhaps the most important point of this chapter is to demonstrate that the effects of cumulative errors must be recognized and optimized. This book was written specifically to call attention to the various parameters that must be checked before attempting any significant error analysis. It is hoped that in this way some of the common pitfalls in system design can be minimized.

9.7. SYSTEM TESTS

In Section 9.2 we considered the problem of interpreting vendor specifications so that a valid comparison between "apples and apples" could be made. An equally important problem is the design of a testing method that closely follows the final usage.

Most system engineers are thoroughly schooled in reading government specifications where each paragraph must be scrutinized as an entity. Very often the philosophy has been to meet the specification even if its end-product validity is seriously in question. A corollary is that the low bidder usually gets first preference when awards are made. In commercial contracts, on the other hand, the contractor, once the contract is awarded to him, is considered to be the expert on the proposed project; it is his responsibility to develop realistic test methods within the original framework of the contract. It is therefore necessary to study the customer's operation so that test methods compatible with his equipment and personnel can be developed. If it is impractical for him to conduct complete tests, some reasonably safe approximation or qualitative tests should be made. For example, an optical tachometer used in the research facilities of a large jet engine manufacturer could reasonably be expected to be checked by well-trained technical specialists; they could probably perform any test that the original manufacturer desired. Conversely, a flowmeter used in a food processing plant

would probably be cared for by less technically sophisticated personnel. It would be unreasonable to expect them to be capable of calibrating the flowmeter, but some qualitative indication of an actual or projected failure should be provided.

The philosophy of system tests should be based on practicality rather than on laboratory conditions. For example, the flowmeter previously discussed may be accurate to within 0.1% under ideal conditions; the important question is what happens to it when the system is subjected to cavitation during a start-up operation. How can the start-up be synthesized in the laboratory and a reasonable life cycle for the flowmeter established? How often should the user check the system under these special conditions? Questions such as these are the responsibility of the systems engineer; complete answers to these points are just as important as the stress analysis on the bearings.

The elements of successful system testing are as follows:

1. Test data that demonstrate the ability of the system to perform satisfactorily in the operating environment.
2. Establishment of criteria for periodic examination by the customer.
3. Details on indications of incipient failure.
4. Test data obtained at the customer's plant with the equipment and personnel that are expected to use the system. This both tests the system and trains the personnel in its proper use.

9.8. SCHEDULING

Scheduling the design and development of an electromechanical system is one of the most demanding tasks that an engineer must perform. The key events can be summarized as follows:

1. Analysis.
2. Preliminary design.
3. Selection of components.
4. Final design.
5. Construction of breadboards or prototypes.
6. Debugging, optimization of design, and testing.
7. Preparation of production drawings and fixtures.
8. Production phase.

Producing a PERT chart to summarize these events becomes a problem only if an overzealous administrator requires a large number of charts for each event. Excessive use of PERT leads to so complex a system that it

becomes an end in itself and discourages the cooperation of the technical staff. The real problem is establishing realistic dates. Estimates of engineering time are traditionally too low, insufficient allowances being made for debugging and unexpected problems. The stress analysis, thermal analysis, kinematic analysis, and circuit designs that are "routine" rarely go as planned. Allowance for exotic designs generally are more realistic because there is less tendency to make optimistic assumptions. Time estimates should be based on records on past jobs; if these are not available, use the time estimated by the man who will probably do the job and add to it a healthy safety factor. Remember, however, that too large a safety factor will stretch out the time required to do the job; you may not get the award then.

Determining the delivery time of components from outside vendors is still an art rather than a science. For a standard part that is made by more than one company estimating the delivery time is relatively simple, provided that such common pitfalls as finding its distributor, getting the purchase order released, and passing incoming inspection are accounted for in the estimate. Such delay-causing factors belong in the "iceberg" category in that 90% of the problems are below the surface. When a nonstandard or new component is required, delivery time is much more difficult to predict. A vendor may be able to quote a low price and promise quick delivery merely because he is unacquainted with the design subtleties of the product. High prices and long delivery times may be equally noninformative. Perhaps the best criterion is the vendor's job record on similar work done for a competitive company. Unlike in other fields, competing engineers usually exchange this type of information for mutually beneficial reasons. When such information is not available, an inspection of the vendor's plant is imperative. A preliminary step is to secure data on his financial stability.

The plant inspection should start with the machinery. Are the necessary lathes, milling machines, drill presses, chucks, and screw machines actually on the premises, or will the vendor subcontract the work? Is the equipment properly installed and in good working order? Even the best equipment needs periodic maintenance to guarantee the condition of the bearings and associated parts. If the intended job involves high-production machining, the availability of automatic equipment should be checked. If conventional chassis and discrete components are to be assembled, chassis-forming equipment should be available. The various production aids used for wiring should be in evidence—harness-forming boards, small precision soldering irons, or possibly microsoldering machines. If the work involves printed circuits, some dip soldering equipment should be available. In summary, the tools to do the required work should be immediately available for inspection and in good working order or the buyer may find

himself chasing down small basement operations when the finished parts are critically needed.

Another basic question is whether the vendor has the personnel to operate the equipment. The determination of this is a rather subjective process. An experienced machinist can spot another craftsman by the way he handles his tools, without having to inspect the final product. The same is true of wiremen and assemblers. For this reason an individual experienced in shop work must be present in any plant inspection. If the project involves only assembly work and the personnel must be trained for the job, add an extra safety factor to the delivery time promised by the vendor.

Does the vendor have the facilities for producing the various jigs, fixtures, and assembly aids required for the operation? Is there a supporting staff to design and debug them?

The inspection or quality-control department is the most important part of the operation. The department need not be elaborate to be effective. The effectiveness of their personnel depends on how closely they monitor the work in process and what plans they have for inspecting the final product before it reaches you. How many parts are inspected—1% or 100%? What controls are imposed on incoming materials? Do they have the scope, secondary standards, bridges, hardness testers, and optical comparators that the work requires? Are there written records to support their claims that the department actually functions in the expected manner? A Q.C. department that never rejects anything will not be of any value to you. The physical appearance of the inspection room is usually a key to the role of the group within the company. Other things being equal, a dirty, untidy inspection area is indicative of a group that is bypassed by the boss when shipments must be delivered at the end of the month. In the electro-mechanical field one of the key items is plated parts. Does the Q.C. department have anyone who knows the difference between good and bad plating? If there are other critical processes involved, does the vendor have an inspector knowledgeable in this field?

When the answers to all these questions are reasonably positive, the final step is to discuss again the fine points of the contract with the vendor, to avoid mutual or unilateral misunderstandings.

Unfortunately, many firms have long been in the practice of accepting a contract with full knowledge that the end product cannot be produced without some key change or that it cannot be produced on time. The reason is simple: if the customer supplies mechanical prints containing parts that cannot be assembled (such as an oversized shaft and mating bushing), the customer must pay for the error, the vendor's profit is increased, and his responsibility for the delivery date is relaxed. This is sometimes referred to as a "change of scope" in the original contract. The systems engineer is

particularly vulnerable to this practice when the product is being rushed from the experimental to the production stage; under these conditions there is rarely time to check the drawings exhaustively. When the error involves a long lead time item, it can be disastrous. Sometimes the vendor succeeds in convincing the systems engineer that he can fulfill the contract in record time, even though he knows that he has little chance of doing so. His reasoning is that once the contract has been signed, the customer will prefer to carry it through regardless of time delays rather than cancel and start anew with a new contractor, which will require even more time. The only way of preventing this is to periodically monitor the progress of the vendor and to impose monetary penalties for poor performance.

All the points mentioned above must be carefully evaluated before any successful scheduling can be completed.

9.9. CONCLUSION

The information and philosophy in this book are designed to provide the engineer with a practical guide to the real world of engineering. The cardinal point is that the data in the literature on components should be evaluated with healthy skepticism. This skepticism will ultimately provide the basis of better future components and systems.

REFERENCES

1. I. Cochin *Analysis and Design of the Gyroscope for Inertial Guidance*, John Wiley & Sons, Inc., New York, 1963.
2. V. Del Toro and S. R. Parker, *Principles of Control Systems Engineering*, McGraw-Hill Book Company, New York, 1960.
3. H. Chestnut, *Systems Engineering Methods*, John Wiley & Sons, Inc., New York, 1967.

Appendix A Encoder Definitions

Accuracy. The maximum positional difference between the input to an encoder and the position indicated by its output; includes both deviation from theoretical code transition positions and quantizing uncertainty caused by converting from a scale having an infinite number of points to a digital representation containing a finite number of points. *See* Resolution.

Ambiguity. Inherent error caused by multiple bit changes at code transition positions, which is eliminated by various scanning techniques. *See* Scan.

Alphanumeric code. *See* Code.

BCD. *See* Code, Binary coded decimal.

Binary. Pertaining to number systems based upon the Radix 2.

Binary coded decimal code. *See* Code.

Binary word. A related grouping of "1" and "0" having meaning assigned by definition, or weighted numerical value in the natural binary system of numbers.

Bit. One element of a two state digital code word having a value of 1 or 0, "on" or "off," etc.

Bit width. The angular increment of input position defined by either the true "1" or false "0" value of a bit. For example: the bit width of the least significant digit of a Gray code comprises two quanta.

Bi-quinary code. *See* Code.

Block V-scan. *See* Scan.

Code. A group of binary bits uniquely arranged which identifies an encoder input position.

Binary coded decimal (BCD). Binary words which translate to decimal numbers.

Bi-quinary. A six-bit code word comprising two code groups: a one-bit and a five-bit group, which translates to a decimal number.

Cyclic decimal code. A four-bit binary code word in which only one digit changes state between any two sequential code words, and which translates to decimal numbers; categorized as one of a group of unit-distance codes.

Decimal. A 10-bit binary word in which there is a one-to-one correspondence between a bit and a decimal digit.

8421 BCD. A binary coded four-bit word in which the first ten natural binary code words represent the decimal values of 0 through 9; 8421 equaling the values of 2^3 , 2^2 , 2^1 , and 2^0 in that order.

Excess-3 BCD (XS-3). A variation of 8421 BCD code in which the natural binary sequence of values from 3 through 12 respectively represent the decimal numbers 0 through 9; used for convenience in forming nine's complements.

Gray (reflected binary). A unit distance code obtained by a reflection of each bit in the natural binary code.

Gray coded excess-3 BCD. Gray coded versions of excess-3 BCD code words which translate to decimal numbers. Categorized as one of a group of cyclic decimal codes.

Incremental. An arrangement of 2-bits phased 90° (electrical) apart, from which direction of rotation can be sensed; repeated sequentially and summed algebraically in an external counter.

Incremental code, non direction sensing. A series of pulses representing equal intervals of angle or equal increments of linear motion.

Natural binary. A number system to the base (Radix) 2, in which the "1" and "0" have weighted value in accordance with their relative position in the binary word.

Octal, binary coded. A binary coded number system to the base (Radix) 8, in which the natural binary values of 0 through 7 are used to represent octal digits with values from 0 to 7 within each octave.

Unit distance. Any sequence of code words in which only one bit changes state between any two adjacent words and in which the respective values of the code words are prescribed by definition.

Command. *See* Interrogation.

Complement. *See* ONE's complement, NINE's complement.

Counts per turn. *See* Resolution.

Cyclic decimal code. *See* Code.

Decimal encoder. An encoder having ten output lines for each decade of decimal numbers; one line representing each digit from 0 to 9.

Digit. A symbol used to represent integers smaller than the Radix, e.g., 0 and 1 in binary notation (synonymous with bits), 0 through 7 in octal notation, 0 through 9 in decimal notation.

Diode U-scan. *See* Scan.

8421 BCD code. *See* Code.

Excess-3 BCD code. *See* Code.

Gray Code. *See* Code.

Gray coded excess-3 code. *See* Code.

Incremental codes. *See* Code.

Interface

Electrical. Electrical interconnection between system elements.

Mechanical. Mechanical mounting and interconnections between system elements.

Interrogation

Pulse. Periodic synchronous or asynchronous electrical activation and observance of an encoder's shaft position.

Continuous. Constant electrical activation and observance of an encoder's reported shaft position.

Natural binary code. *See* Code.

NINE's complement. That decimal digit which, when added to any other decimal digit, equals 9.

Noise. *See* Figure 1.

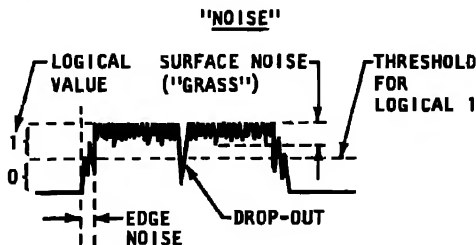


Figure 1. Noise.

Octal code. *See* Code.

Octave. A term related to octal code, used in the same sense as "decade" relative to decimal and "bit" relative to natural binary.

ONE's complement. That binary bit which, when added to 1 or 0, equals 1; synonymous with the inverse binary state of any given bit.

Parallel output. Simultaneous availability of two or more bits, channels, or digits.

Parity bit. An additional bit added to a code word to make the sum of the word bits always "even parity" or "odd parity."

Quantize. To subdivide the range of a variable into a finite number of non-overlapping intervals, each of which is assigned a specific code identity.

Quantizing deviation. See Figure 2.

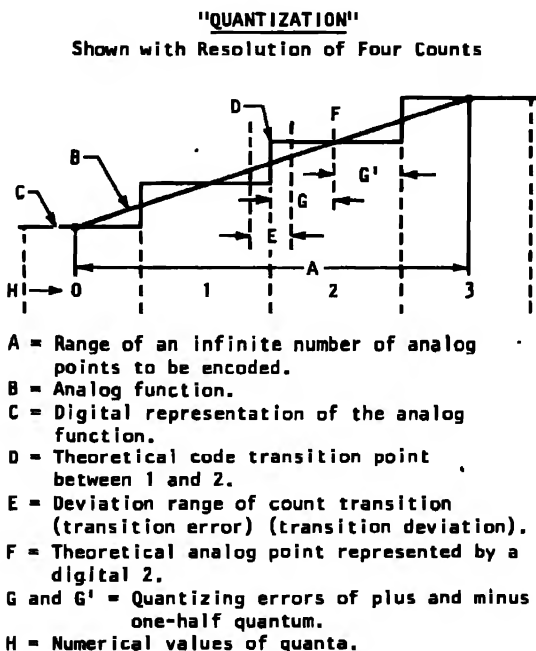


Figure 2. Quantization, shown with a resolution of four counts.

Quantum. The increment of input position defining a specific code identity.

Resolution. A measure of code position density. Can be defined as the number of quanta or counts per input shaft revolution or as a power of 2 (7 bits in a 2^7 -bit-per-turn encoder) or as a quantum. Thus a 7 bit/turn natural binary encoder could be said to have a resolution of 128 counts per turn, 7 bits, or $2^\circ 48' 45''$ ($\frac{360}{128}$).

Self decode. A term used to indicate that all or a part of the scanning is done internal to the encoder.

Self decoding U-scan. *See* Self decode.

Self decoding V-scan. *See* Self decode.

Self selecting U-scan. *See* Self decode.

Self selecting V-scan. *See* Self decode.

Scan. To readout by a method which eliminates data ambiguity.

Block V-scan. A combination of U-scan and V-scan required for BCD codes other than unit-distance codes, for avoidance of ambiguity. U-scanned binary code words are V-scanned in blocks of four bits each or less.

Diode U-scan. A variation of U-scan.

U-scan. An encoder interrogation method in which one of two sets of sensors (2 per bit), are selected in parallel. The set of sensors readout are selected by the state of a selector bit. Named for the geometric position of the sensors which resembles a "U."

U-scan, self decoding. Synonymous with U-scan.

V-scan. An encoder interrogation method in which one of two sensors per bit are selected to be read out by the state of the next lower order (lesser significant) binary bit. Interrogation is serial in effect. The term V-scan is derived from the geometric placement of the sensors which resembles a "V."

V-scan, self decoding. A term used to indicate that all or part of the scanning is done internal to the encoder.

V-scan, self decoding. Synonymous with self decoding V-scan.

Serial output. Sequential availability of two or more bits, channels, or digits.

Transition deviation. *See* Figure 2. The difference between the theoretical encoder position reading and the actual encoder position reading.

Unit distance code. *See* Code.

U-scan. *See* Scan.

V-scan. *See* Scan.

Word. *See* Binary word.

Weighted value. The numerical value assigned to any single bit as a function of its position in the code word.

Appendix B Potentiometer Definitions

The definitions used here are those recommended by the Precision Potentiometer Manufacturers Association. Only those pertinent to the discussion in this article are included. A full set of these recommended definitions may be obtained from: PPMA, 3525 Peterson Rd., Chicago 45, Illinois.

List of Symbols

C = conformity	R_T = total resistance
CT = center tap	R_e = end resistance
CW = clockwise	TC = temperature coefficient of resistance
CCW = counterclockwise	θ = shaft position
E = total applied voltage	θ_T = theoretical electrical travel
e = output voltage	θ_A = actual electrical travel

INPUT AND OUTPUT

Total applied voltage. The voltage applied between input terminals.

Output ratio. The ratio of the output voltage to the total applied voltage.

Total variable output. The difference between the maximum and minimum output ratios.

Loading error. The difference between the actual output ratios at a specified shaft position with some specified wiper load in place of an infinite wiper load.

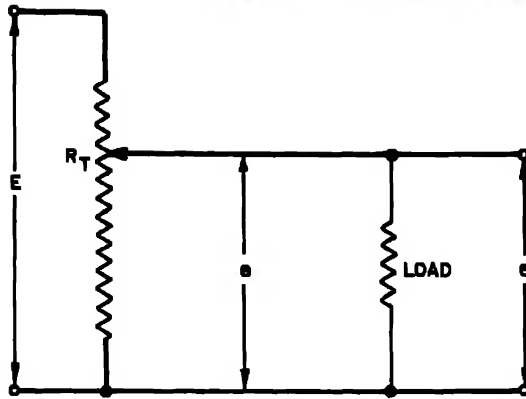


Figure 1. Potentiometer schematic.

RESISTANCE

Total resistance. The resistance between the input terminals with the shaft positioned so as to give a maximum resistance value.

Minimum resistance. The resistance measured between the wiper terminal and any terminal with the shaft positioned to give a minimum value.

End resistance. The resistance measured between the wiper terminal and an end terminal with the shaft positioned to give a minimum resistance value.

ROTATION AND TRAVEL

Direction of rotation. Shaft rotation is defined as clockwise (CW) or counterclockwise (CCW) when viewing the specified mounting end of the potentiometer.

Direction of translation. Shaft translation is defined as “extending” or “retracting” when viewing the specified end of the potentiometer. (Applies to translatory potentiometers only.)

Total mechanical travel. The total travel of the shaft between integral stops, under specified stop torque. In potentiometers without stops, the mechanical travel is continuous.

Mechanical overtravel. The shaft travel between each end point and its adjacent mechanical stop.

Theoretical electrical travel. The shaft travel over which the theoretical function characteristic extends.

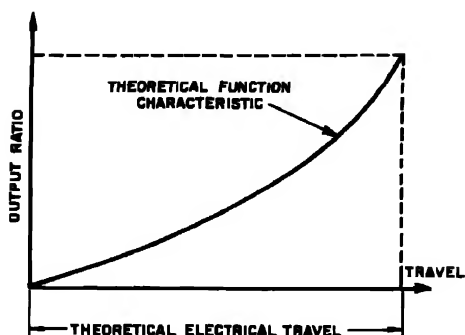


Figure 2. Potentiometer output ratio versus travel.

Electrical overtravel. The shaft travel over which there is continuity between the wiper terminal and the resistance element beyond each end of the actual electrical travel. (In cases where absolute linearity or absolute conformity is specified, "theoretical electrical travel" shall be substituted for "actual electrical travel" in this definition.)

Electrical continuity travel. The total travel of the shaft over which electrical continuity is maintained.

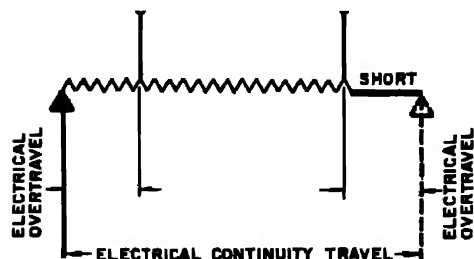


Figure 3. Potentiometer overtravel.

Actual electrical travel. The total travel of the shaft between end points.

END POINT AND TAP

End point. The shaft position immediately before (for practical measurement of this point as related to mechanical position or travel only, this point must be taken at the first measurable change of output voltage) the first measurable change of actual output ratio is observed as the shaft moves the wiper in a specified direction from the overtravel region on to the region of actual electrical travel. The other end point is the shaft position at which the final (maximum or minimum) actual output ratio first occurs while the shaft is still moving in the same direction.

End voltage. The voltage between the wiper terminal and an end terminal when the shaft is positioned at the end point. It is usually expressed as a percentage of the total applied voltage.

Jump-off voltage. The first measurable voltage change as the shaft moves the wiper from the overtravel region on to the actual electrical travel region. It is usually expressed as a percentage of the total applied voltage.

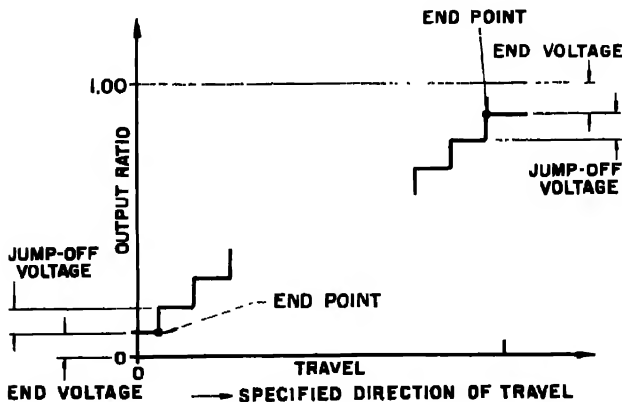


Figure 4. Potentiometer jump-off voltage.

Tap. A fixed electrical connection made to the resistance element.

Tap location. The position of a tap from some reference point. (This is commonly expressed in terms of resistance, voltage ratio, and/or shaft position. When a shaft position is specified, the tap position is measured at the center of the effective tap width.)

Effective tap width. The travel of the shaft during which the voltage at the wiper terminal and the tap terminal are essentially the same as the wiper is moved past the tap in one direction.

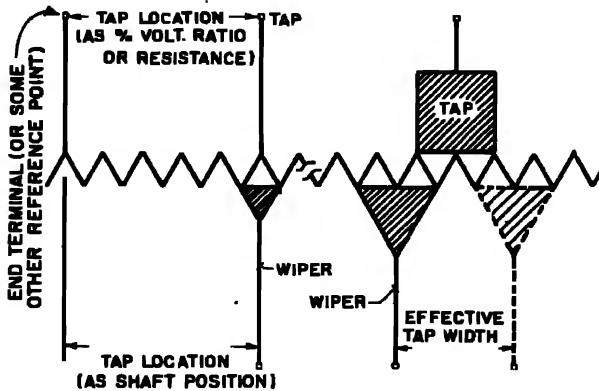


Figure 5. Potentiometer tap location.

RESOLUTION

Resolution. The measure of the sensitivity to which the output ratio of the potentiometer may be set.

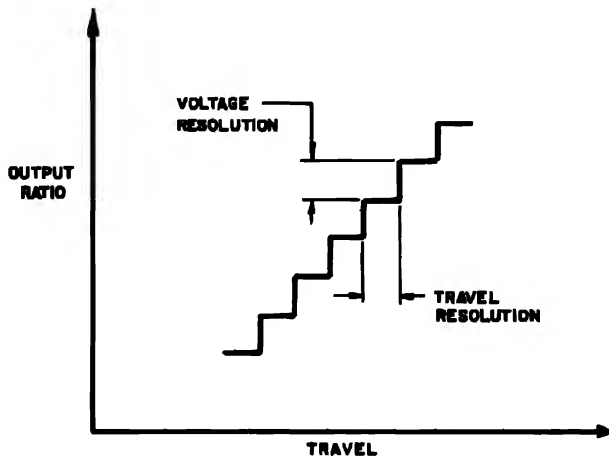


Figure 6. Potentiometer resolution.

Theoretical resolution. Used in wirewound linear potentiometers only, it is the reciprocal of the numbers of turns of the resistance winding in the actual electrical travel, and is expressed as a percentage:

$$N = \text{total number of resistance wire turns}$$

$$100/N = \text{theoretical resolution in percent}$$

Travel resolution. The maximum value of shaft travel in one direction per incremental voltage step in any specified portion of the resistance element.

Voltage resolution. The maximum incremental change in output ratio with shaft travel in one direction in any specified portion of the resistance element.

PHASING, TRACKING, AND INDEXING

Index point. A point of reference fixing the relationship between a specified shaft position and the output ratio. It is used to establish a shaft position reference.

Phasing. The relative alignment of the cups of a gang with respect to the position of the wipers on their respective electrical elements.

Phasing point. A point of reference fixing the relationship between the shaft position and the output ratio, or between the shaft position and a tap location, for each electrical element in a gang potentiometer.

Simultaneous conformity phasing. The alignment of the electrical elements of a gang potentiometer so that the output ratios fall within their respective conformity limits over the theoretical electrical travel, using a common shaft position reference. This definition applies only to potentiometers with absolute conformity or absolute linearity function specifications.

Voltage tracking error. The difference at any shaft position between the output ratios of any two commonly actuated similar electrical elements expressed as a percentage of the single total voltage applied to them.

CONFORMITY AND LINEARITY

Function characteristic. The relationship between the output ratio and the shaft position. Mathematically:

$$\frac{e}{E} = f(\theta)$$

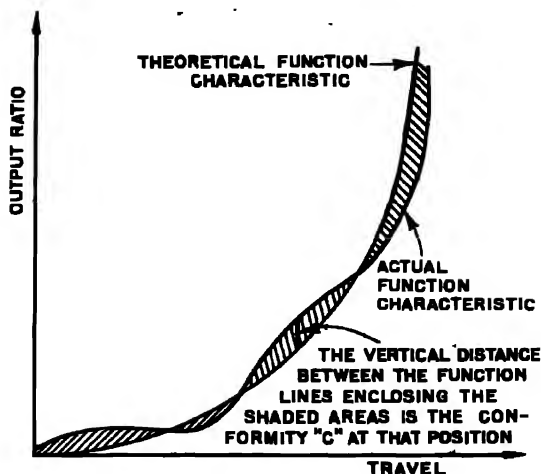


Figure 7. Potentiometer conformity.

Conformity. The fidelity of the relationship between the actual function characteristic and the theoretical function characteristic. Mathematically:

$$\frac{e}{E} = f(\theta) \pm C$$

Absolute conformity. The maximum deviation, expressed as a percent of the total applied voltage, of the actual function characteristic from a theoretical function characteristic extending between the specified output ratios which are separated by the theoretical electrical travel. An "index point" on the actual output is required. Mathematically:

$$\frac{e}{E} = f\left(\frac{\theta}{\theta_T}\right) \pm C$$

$$0 \leq \theta \leq \theta_T$$

Linearity. A specific type of conformity where the theoretical function characteristic is a straight line. Mathematically:

$$\frac{e}{E} = f(\theta) \pm C = A(\theta) + B \pm C$$

where A = given slope

B = given intercept at $\theta = 0$

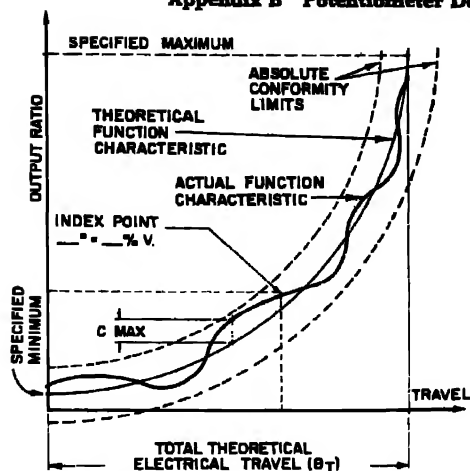


Figure 8. Potentiometer absolute conformity limits.

Absolute linearity. The maximum deviation, expressed as a percent of the total applied voltage, of the actual function characteristic from a straight reference line drawn through the specified minimum and maximum output ratios which are separated by the theoretical electrical travel. Unless otherwise specified, minimum and maximum are respectively zero and 100% total applied voltage. An "index point" on the actual output is required. Mathematically:

$$\frac{e}{E} = A \left(\frac{\theta}{\theta_T} \right) + B \pm C$$

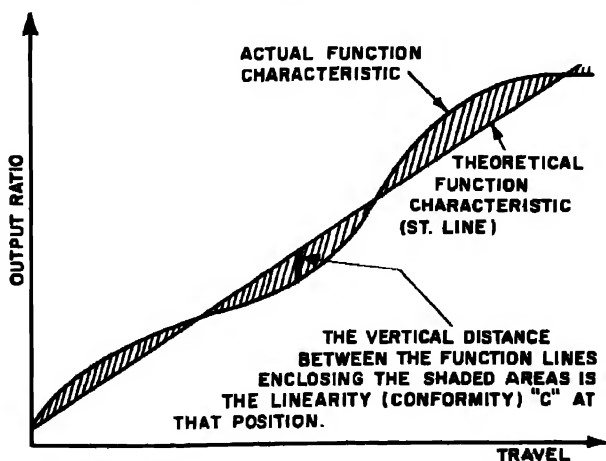


Figure 9. Potentiometer absolute linearity.

where A = given slope

B = given intercept at $\Theta = 0$

unless otherwise specified, $A = 1$ and $B = 0$.

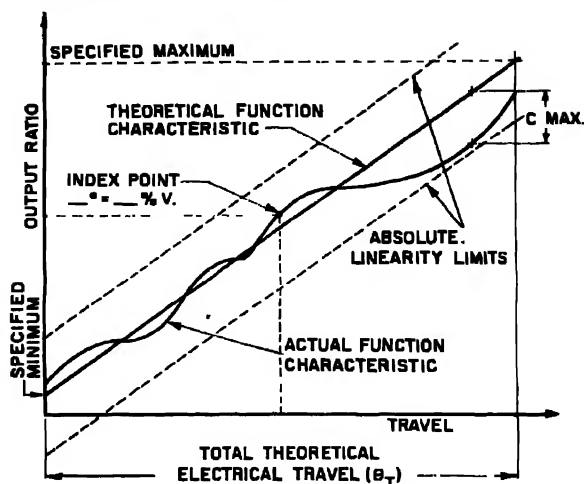


Figure 10. Potentiometer linearity defined.

Terminal based linearity. The maximum deviation, expressed as a percent of the total applied voltage, of the actual function characteristic from a straight reference line drawn through the specified minimum and maximum output voltage ratios which are separated by the actual electrical travel. Unless otherwise specified, minimum and maximum output ratios are respectively zero and 100% of total applied voltage. Mathematically:

$$\frac{e}{E} = A \frac{\theta}{\theta_A} + B \pm C$$

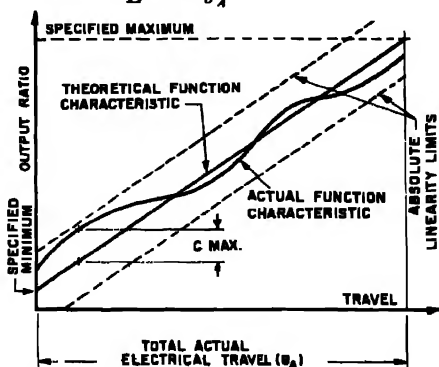


Figure 11. Potentiometer terminal based linearity.

where A = given slope

B = given intercept at $\Theta = 0$

unless otherwise specified, $A = 1$ and $B = 0$.

Zero-based linearity. The maximum deviation, expressed as a percent of total applied voltage, of the actual function characteristic from a straight reference line drawn through the specified minimum output ratio, extended over the actual electrical travel, and rotated to minimize the maximum deviations. Any specified end voltage requirement limits the rotation of the reference line. Unless otherwise specified, the specified minimum output ratio will be zero. Mathematically:

$$\frac{e}{E} = P \frac{\theta}{\theta_A} + B \pm C$$

where P is unspecified slope limited by the end voltage requirements, at the maximum output ratio end. Unless otherwise specified, $B = 0$.

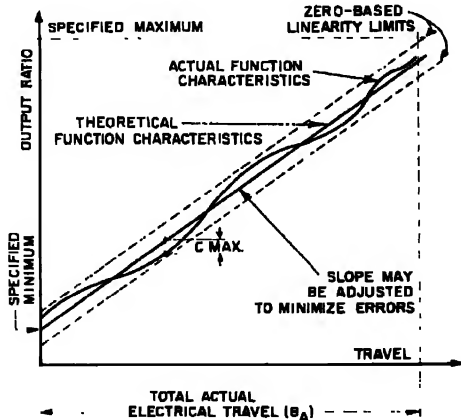


Figure 12. Potentiometer zero based linearity.

Independent linearity (best straight line). The maximum deviation, expressed as a percent of the total applied voltage, of the actual function characteristic from a straight reference line with its slope and position chosen to minimize the maximum deviations over the actual electrical travel, or any specified portion thereof. (*Note.* End voltage requirements when specified will limit the slope and position of the reference line.) Mathematically:

$$\frac{e}{E} = P \frac{\theta}{\theta_A} + Q \pm C$$

where P = unspecified slope

Q = unspecified intercept at $\Theta = 0$

and both are chosen to minimize C but are limited by the end voltage requirements.

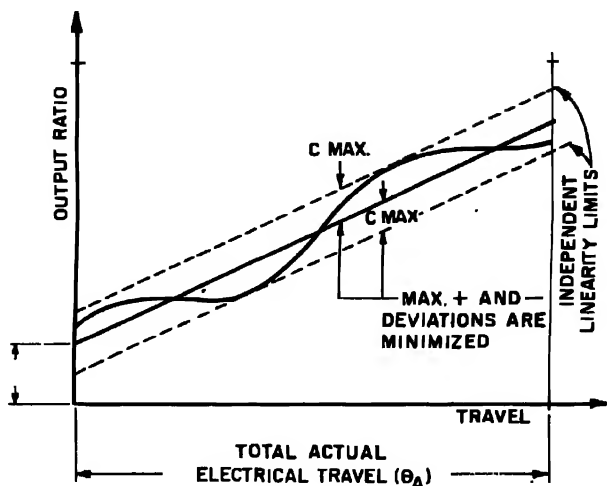


Figure 13. Potentiometer independent linearity.

Tolerance limits; alternate methods. There are three basic methods:

1. *Constant limits.* Taken as a percentage of the total applied voltage.
2. *Proportional limits.* Taken as a percentage of the theoretical output voltage ratio.
3. *Modified proportional limits.* Any combination of the first two methods.

All PPMA definitions employ method 1 for stating tolerance limits. Note that proportional limits may become impossibly restrictive in the vicinity of zero output, and should be modified in such cases to provide a practical tolerance in that region.

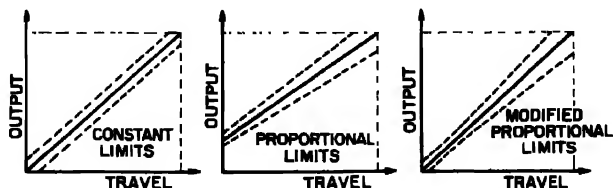


Figure 14. Potentiometer tolerance limits.

MECHANICAL CHARACTERISTICS

Backlash. The maximum difference in shaft position that occurs when the shaft is moved to the same actual output point ratio from opposite directions. Resolution effects must be excluded from this measurement.

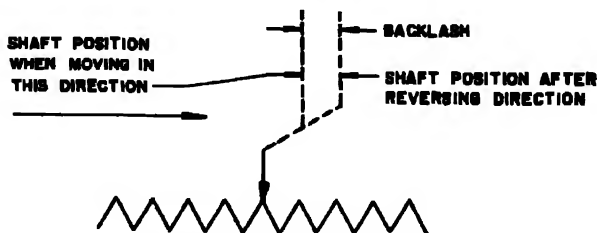


Figure 15. Potentiometer backlash.

Note. The following definitions apply to rotary potentiometers only.

Shaft runout. The eccentricity of the shaft diameter with respect to the rotational axis of the shaft, expressed in inches, and measured at a specified distance from the mounting face when the body of the potentiometer is held and the shaft rotated while a specified load is applied radially to the shaft.

Lateral runout. The perpendicularity of the mounting surface with respect to the rotational axis of the shaft, expressed in inches and measured on the mounting surface at a specified distance from the axis of rotation when the shaft is held and the body of the potentiometer is rotated while specified loads are applied radially and axially to the body of the pot.

Pilot diameter runout. The eccentricity of the pilot diameter with respect to the rotational axis of the shaft expressed in inches and measured on the pilot diameter when the shaft is held and the body of the potentiometer is rotated while a specified load is applied radially to the body of the pot.

Shaft radial play. The total radial excursion of the shaft, expressed in inches, and measured at a specified distance from the face of the unit, with a specified radial load applied alternately in opposite directions at a specified point.

Shaft end play. The total axial excursion of the shaft, expressed in inches, and measured at the end of the shaft with a specified axial load applied alternately in opposite directions.

Starting torque. The maximum moment in the clockwise and counter-clockwise direction required to initiate shaft rotation anywhere in the total mechanical travel.

Running torque. The maximum moment in the clockwise and counter-clockwise direction required to sustain shaft rotation at a specified speed throughout the total mechanical travel.

Moment of inertia. The mass moment of inertia of the rotating element of the potentiometer about its rotational axis. (Includes shaft and connected rotating members.)

STOP STRENGTH

Static stop strength. The maximum load that can be applied to the shaft at each stop without a permanent change of the stop position greater than specified.

Dynamic stop strength. The inertia load, at a specified shaft velocity and a specified number of impacts, that can be applied to the shaft at each stop without a permanent change of the stop position greater than specified.

ELECTRICAL CHARACTERISTICS

Noise. Any spurious variation in the electrical output not present in the input, defined quantitatively in terms of an equivalent parasitic, transient, resistance in ohms, appearing between the contact and the resistance element when the shaft is rotated or translated. The equivalent noise resistance is defined independently of the resolution, the functional characteristics, and the total travel. The magnitude of the equivalent noise resistance is the maximum departure from a specified reference line. The wiper of the potentiometer is required to be excited by a specified current and moved at a specified speed.

Life. The life expectancy of a potentiometer is the number of shaft revolutions or translations obtainable under specific operating conditions and within specified allowable degradations of specific characteristics.

Temperature coefficient of resistance. The unit change in resistance per degree centigrade change from a reference temperature, and expressed in parts per million per degree centigrade as follows:

$$\text{T.C.} = \frac{R_2 - R_1}{R_1(T_2 - T_1)} \times 10^6$$

where R_1 = resistance at reference temperature in ohms

R_2 = resistance at test temperature in ohms

T_1 = reference temperature in degrees centigrade

T_2 = test temperature in degrees centigrade.

Power rating. The maximum power that a potentiometer can dissipate under specified conditions while meeting specified performance requirements.

Dielectric strength. Ability to withstand a specified potential of a given characteristic between the terminals of each cup and the exposed conducting surfaces of the potentiometer, or between the terminals of each cup and the terminals of every other cup in the gang under prescribed conditions without exceeding a specified leakage current value.

Insulation resistance. The resistance to a specified impressed DC voltage between the terminals of each cup and the exposed conducting surfaces of the potentiometer, or between the terminals of each cup and the terminals of every other cup in the gang under prescribed conditions.

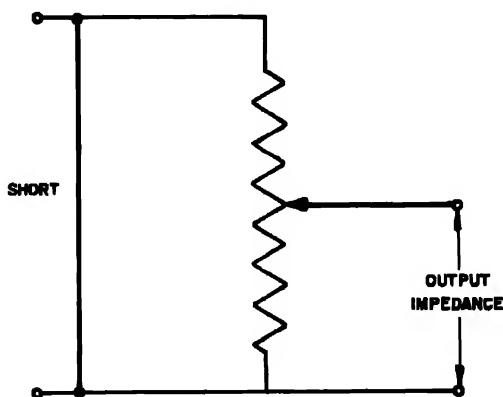


Figure 16. Potentiometer insulation resistance.

Voltage short. A segment of the resistance element over which the output ratio remains constant within specified limits as the wiper traverses the segment.

Resistance short. A segment of the resistance element over which the resistance between the wiper and a specified terminal remains constant within specified limits as the wiper traverses the segment.

AC CHARACTERISTICS

Total input impedance. The impedance between the two input terminals with open circuit between output terminals, and measured at a specified voltage and frequency with the shaft positioned to give a maximum value.

Output impedance. Maximum impedance between slider and either end terminal with the input shorted, and measured at a specified voltage and frequency.

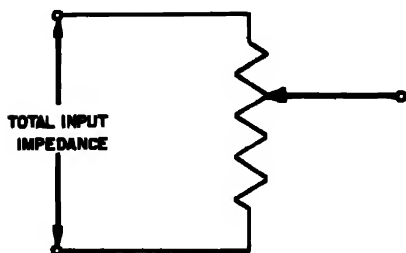


Figure 17. Potentiometer output impedance.

Phase shift. The maximum phase difference measured in degrees between the sinusoidal input and output voltages measured at a specified input voltage and frequency.

Quadrature voltage. The maximum value of that portion of the output voltage which is $\pm 90^\circ$ out of time phase with the input voltage, expressed as volts per volt applied, measured at a specified input voltage and frequency.

SOURCE. *Electromechanical Design*, January 1964, Benwill Publishing Company, Brookline, Mass.

Appendix C Bearing Terminology

Abrasion. Wear in a bearing material and shaft caused by foreign hard particles; usually resulting from repeated scoring.

Axial clearance. The total axial movement of the unclamped ring when a specified load moves first in one direction and then in the other direction.

Back-to-back mounting. Mounting arrangement in which two single-row ball bearings, preloaded against each other, have contact-angle lines which diverge toward the axis of rotation.

Ball complement. The number of balls contained in a bearing.

Ball groove. The raceway in a ball bearing.

Ball-indented surface. A strip-type plain bearing surface that contains ball-shaped recessions for use as small oil or graphite reservoirs.

Ball pocket. The internal surface of a ball bearing cage that contacts the periphery of the ball.

Ball-riding retainer. A ball cage, supported concentric to inner and outer rings by its contact with the bearing balls.

Bearings. A support or guide which positions a moving part with respect to the other parts of a mechanism.

Bearing characteristic number. A dimensionless number which evaluates the performance of a plain bearing. It relates load, clearance, diameter, speed and viscosity.

Bleeding. The tendency of a liquid component to separate from a liquid-solid or liquid-semisolid mixture; as an oil from a grease.

Block grease. A grease of high dropping point which, under normal temperatures, is firm to the touch and can be handled in block or stick form.

Block penetration. The penetration (at 77F) of a grease that is sufficiently hard to hold its shape. (ASTM Designation D217-52.)

Boundary dimensions. The bore, OD and width of a bearing.

Boundary lubrication. A lubrication condition in which the properties of the surfaces and the properties of the lubricant other than viscosity determine friction between two surfaces in relative motion.

Brinell. A small permanent deformation in a raceway or rolling element which is caused by a static overload.

Brown oils. Fatty oils (such as rapeseed, whale, or various fish oils) artificially thickened by blowing a current of air through them.

Cage. A device partly surrounding the rolling elements and traveling with them. It spaces the rolling elements in proper relation to each other. Also called retainer or separator.

Cartridge bearing. Scaled, grease-lubricated, rolling element bearings with extra-wide rings.

Centipoise. The unit of absolute viscosity. 1 centipoise equals 0.01 poise.

Centistoke. The unit of kinematic viscosity. 1 centistoke equals 0.01 stoke.

Centralized lubrication. A lubrication system which supplies the lubricant or lubricants for the bearings of a machine or group of machines.

Circumferential clearance. The total clearance between rolling elements measured along their pitch circle.

Clearance bearing. A journal bearing in which the bearing surface radius exceeds the journal surface radius.

Clearance ratio. The diametral clearance divided by the journal diameter.

Compatability. A measure of the anti-weld or antiscoring characteristics of a plain bearing material when operated with a given mating material.

Composite bearing. A plain bearing made with two or more different bearing materials which give improved properties.

Cone. The inner ring and race of a tapered-roller bearing.

Conformability. A quality of a plain bearing material which permits the material to adjust itself to deflections without developing a high-temperature condition.

Contact angle. The angle between a plane perpendicular to the axis of rotation and a line connecting the points of tangency of the rolling elements with the inner and outer raceways.

Contact area. The projection of the contact surface between a rolling element and a raceway on a plane perpendicular to the contact angle.

Contact ellipse. The contact area of a ball bearing.

Crowned roller. A roller having a slight curvature to its surface of revolution, thus preventing roller end-loading and permitting better stress distribution.

Cup. The outer ring and race of a tapered-roller bearing; also, the outer ring and race of certain needle bearings.

Deformability. That quality of a plain bearing material which permits it to yield to deformation under operation without causing failure.

Detergent. An additive or a compound lubricant which keeps insoluble matter in suspension, thus preventing its deposition where it would be harmful. A detergent may also redisperse deposits already formed.

Diametral clearance. The bearing bore diameter minus the journal diameter.

Drawn cup. A thinshell outer ring and race on certain needle bearings.

Drop-feed lubrication. A lubrication system which supplies drops of lubricant to the bearing surfaces at regular intervals.

Dropping point. The temperature at which grease passes from a semisolid to a liquid state under specified test conditions (ASTM Designation D 566-42).

Duplexed bearings. Matched sets of bearings usually preloaded.

Dynamic load capacity (radial). The radial load that a bearing can support for a rating life of one million revolutions (500 hr at $33\frac{1}{3}$ rpm).

Eccentricity. The radial displacement of the journal center from the center of the bore.

Eccentricity ratio. Equals the eccentricity divided by the radial clearance.

Effective spread. The distance between the intersection of the contact-angle lines and the axis of rotation of opposed mounted bearings. It is used in accurate bearing load calculations for opposed bearing mounting.

Elliptical bearing. A plain bearing with an approximately elliptical bore formed from two sections of a cylinder. It is more stable when considering oil whip in high-speed bearings than a cylindrical bearing.

Embeddability. A measure of the ability of a plain bearing material to absorb dirt and grit particles.

End cap. A retaining device that holds a bearing in a housing and also forms a lubricant retaining chamber.

Equivalent radial load. The calculated pure radial load having the same effect on bearing life as a given combination of radial and thrust load.

Externally self-aligning bearing. A bearing with a spherical OD on the outer ring which permits alignment of the entire bearing assembly relative to a mating spherical surface.

Extreme-pressure lubricant. Lubricant compounded with certain additives to increase its load-carrying ability.

Face offset. A predetermined stickout of an inner-ring face to the outer-ring face (or vice versa) and which creates a predetermined preload between two bearings when they are clamped together.

Face-to-face. A mounting method which preloads two single-row ball bearings against each other, causing their contact angle lines to converge toward the axis of rotation.

Fatigue failure. Spalling of surface metal caused by repeated stresses under load.

Fatty oil. A fluid composed of fats derived from animal, marine or vegetable origin and which may contain additives.

Fiber grease. A distinctly fibrous-structured grease which (noticeable when a sample of the grease is pulled apart) tends to resist being thrown off gears and out of bearings.

Filler. Any substance (talc, mica or various powders) added to a grease to increase its weight or consistency.

Filling slot. Semicircular notch on the side of the inner and outer ring which permits the assembly of additional rolling elements into a bearing.

Fitted bearing. A partial journal bearing with its bearing surface radius the same as that of the journal surface.

Fixed bearing. A bearing clamped in a housing to prevent axial shaft movement. It locates the shaft and resists thrust in either direction.

Fixed-pad bearing. An axial or radial bearing with fixed pads and contoured surface lands, thus helping to establish a hydrodynamic film.

Flexure. A limited movement type bearing in which the flexure of elastic members rather than rolling or sliding surfaces guide the moving parts.

Floating bearing. A bearing which moves axially with respect to either housing or shaft.

Floating sleeve bearing. A bearing which has one or more sleeves proposed between the stationary bearing and the rotating shaft. The floating sleeve, free to turn, will at ideal conditions rotate at speeds intermediate between those of the shaft and bearing.

Force-feed lubrication. A pressure lubrication system.

Fretting. Wear phenomena taking place between two surfaces having oscillatory relative motion of small amplitude.

Fretting corrosion. Corrosion at the interface between two contacting surfaces, accelerated by vibration and rubbing between the two sufficient to produce localized deformation; sometimes called contact erosion or friction-oxidation.

Full-film lubrication. A continuous film of lubricant separates the journal from the bearing. Normally two kinds are used—hydrodynamic and hydrostatic.

Galling. The damaging of one or both metallic surfaces of a bearing and journal by removal of particles from localized areas during sliding friction.

Graphited surface. An indented plain bearing surface filled with graphite.

Half-frequency whirl. The vibration phenomenon occurring in high-speed, low-load plain bearings. The shaft center rotates around the bearing center at a frequency equal to, or less than one-half the rotational speed of this shaft.

Housing seat. The bore of the housing into which the bearing mounts.

Hydrodynamic lubrication. A lubrication system in which the shape and relative motion of the sliding surfaces cause the formation of a fluid film having sufficient pressure to separate the surfaces.

Hydrostatic lubrication. A lubrication system in which the lubricant, supplied under sufficient external pressure, separates the opposing surfaces by a fluid film.

Inch-series bearing. A bearing having boundary dimensions in integral or fractional-inch figures, rather than metric figures.

Instrument bearing. A precision bearing specifically designed for instrument or miniaturized mechanism applications but which has envelope dimensions exceeding the limits established for miniature bearings.

Internally self-aligning bearing. A bearing in which the curvature of one raceway, swung about a point on the axis of rotation, causes alignment to take place on the rolling elements.

Isoelastic mounting. A bearing mounting with the same yield rate for any load direction—particularly desirable in gyroscope applications.

Journal. That part of a shaft or axle which rotates or angularly oscillates in or against a bearing or about which a bearing rotates or angularly oscillates.

Journal bearing. A sliding-type radial bearing with rotating or oscillatory motion and in conjunction with which a journal operates. In a full journal bearing, the bearing surface extends 360 degrees; in a partial bearing, the bearing surface extends less than 360 degrees.

Journal roller bearing. A rolling-element bearing where the journal of the shaft acts as an inner race.

Lands. Ground cylindrical surfaces of a bearing extending from the raceway shoulder to the bearing faces.

Land-riding retainer. Cage or retainer supported and maintained concentric with the axis of rotation by sliding contact against the lands of either the inner or outer ring.

Length-diameter ratio. A comparison of a plain bearing's length to its diameter.

Life factor. The modifying factor which converts the dynamic load rating to a basis of rating life other than one million revolutions.

Limiting speed. The maximum speed at which a given bearing will safely operate, with proper lubrication.

Linear-motion bearing. A bearing which accommodates axial translation relative to the shaft.

Load track. Dull mat-finished track around the periphery of a raceway indicating the path of the rolling elements around the raceway during service.

Magneto bearing. Separable single-row ball bearing of radial or low-contact-angle type.

Maximum type bearing. Ball bearing having a filling notch (or one counterbored raceway shoulder) which permits assembly of a greater number of balls than possible with conrad construction.

Miniature bearing. Bearing having an OD under $\frac{3}{8}$ in.; or one with $\frac{1}{2}$ in. OD or less and OD-to-bore ratio less than 2.

Mixed-film lubrication. The condition between hydrodynamic lubrication and boundary lubrication. Individual load-carrying pools of self-pressurized lubricant support part of the total load carried by the bearing; the very thin film associated with boundary lubrication supports the remaining part.

Needle bearing. A bearing containing rolling elements which are relatively long compared to their diameter.

Neutral oils. Lubricating oils of low or medium viscosity which, during petroleum distillation, do not undergo treatment with either an acid or an alkali, but are purified by simple filtration.

Nonseparable bearing. Bearing in which you cannot disassemble the inner or outer ring, except by destructive means.

Nutcracker bearing. A plain bearing composed of two cylindrical half-bearings—the lower half remains permanently fixed in housing; the upper half moves freely and is forced downward by hydraulic cylinder. This application of external pressure prevents oil whip.

Oil ring. A loose ring, the inner surface of which rides a shaft or journal, causing the rings to rotate. The ring dips into a lubricant reservoir from which it carries lubricant to the top of the shaft for distribution to a bearing.

Osculation. The degree of conformity between two surfaces in contact.

Overshot bearing. A plain bearing containing a wide circumferential groove in its upper half. The groove carries oil over the top half to reduce power loss.

Overturning moment. The torque applied around an axis perpendicular to the plane that contains the radial and thrust load vectors.

Pad lubrication. A lubrication system that delivers lubricant to a bearing surface by a pad of felt or similar material.

Partial bearing. A slider-type bearing that only partially encloses the journal.

Pitch-circle diameter. The diameter of the circle joining the centers of the rolling elements of a bearing.

Pivot bearing. An axial-load, radial-load type bearing which supports the end of a shaft or pivot (as on the balance wheel of a watch).

Pivoted-pad bearing. An axial or radial load type bearing in which the bearing surface consists of one or more pivoting pads or shoes which help to establish a hydrodynamic film.

Plain bearing. A bearing whose operation depends on sliding motion between two surfaces. Also called a sliding bearing.

Pocket thrust bearing. A slider-thrust bearing consisting of flat plate with recessed areas, or pockets.

Porous bearing. A bearing made from porous material, such as compressed metal powders, the pores acting either as reservoirs for holding, or passages for supplying lubricant.

Pour point. The lowest temperature at which a liquid will flow under specific conditions. (ASTM Designation D 91-52.)

Preload. The self-contained thrust load established between two bearings on the same shaft.

Projected area. The area defined by the diameter times the length of the radial bearing area.

Radial bearing. A bearing which primarily supports a load perpendicular to the shaft axis.

Radial clearance. The total diametric movement of the unclamped ring when a specified load moves first in one direction and then in the other direction. The same as diametral clearance, or bearing bore radius minus journal radius.

Radial load. A load acting in a plane perpendicular to the axis of rotation.

Radial runout. The total radial movement of a point on the stationary outer ring when the inner ring rotates one complete concentric revolution.

Rating life. The number of revolutions (or hours at a given constant speed) that 90% of a group of apparently identical bearings will complete or exceed before the first evidence of fatigue develops.

Retainer. *See* Cage.

Rib. The shoulder directly adjacent to a raceway. It guides the rollers of a roller bearing in their desired path. In cylindrical roller bearings ribs are sometimes used to resist light thrust loads.

Ring lubrication. A lubrication system that uses an oil ring to supply lubricant to the bearing.

Ring-oiled bearing. A bearing lubricated by an oil ring.

Rolling-element bearing. A bearing whose operation depends on the rolling motion of balls or rollers.

Rotation factor. A radial load modifying factor used when the direction of applied load does not rotate with respect to the inner ring.

Scoring. The scratching of a bearing surface caused by relatively hard particles.

Seizure. Bearing failure, usually preceded by some other type of failure, where the torque available in the machine is less than the torque resistance of the damaged bearing.

Self-aligning. The built-in compensation for shaft or housing deflection or misalignment.

Separable. A bearing you can completely or partially separate into its components and then readily reassemble again.

Separator. *See* Cage.

Shield. An annular metal plate stationary with respect to one bearing ring which forms a small running clearance with the other bearing ring. It retains lubricant and prevents entrance of dirt under moderate conditions.

Shoe type bearing. A plain bearing consisting of either the radial-load or axial-load type.

Sleeve bearing. A plain bearing consisting of a band or sleeve that encloses and supports a moving member.

Slushing oil. An oil or grease-like material which, when used on metals, forms a temporary protective coating against rust or corrosion.

Smearing. A special kind of seizing (in its early state) peculiar to rolling-element bearings—the elements slide in the race instead of rolling.

Spalling. The cracking and flaking off of metal particles from a surface; usually caused by fatigue.

Squeeze-film lubrication. The squeeze-film principle applies to plain bearings which operate under reciprocating loads. As the load reverses, the oil replenishes in the clearance space between repeated applications of the load. During load applications there is not enough time for the oil to be completely squeezed out from between the bearing and the journal.

Stand out. The distance the back face of the cone of a tapered roller bearing extends beyond the cup face.

Static load rating. A load which, if exceeded on a non-rotating bearing, produces a total permanent deformation of rolling element and race at the most heavily stressed contact of 0.0001 of the ball or roller diameter.

Step thrust bearing. A thrust bearing with dams or steps which create hydrodynamic action.

Tandem mounting. A mounting method that allows two or more matched ball bearings to share the applied thrust load.

Tapered pad bearing. A fixed pad bearing with its pad surfaces tapered to promote the establishment of a hydrodynamic film.










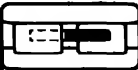


Thrust load. An applied load acting along the axis of rotation.

Underground bearing. A rolling element bearing with underground raceways.

Viscosity, absolute. That property of a fluid, semifluid or semisolid substance which causes it to resist flow. It is defined as the shear stress on a fluid element divided by the rate of shear. The standard unit of absolute viscosity in the English system is the reyn (lb-sec/in²); in the cgs system, the poise (dyne-sec/cm²).

Viscosity index (V.I.). A commonly used measure of a fluid's change of viscosity with temperature. The higher the viscosity index, the smaller the relative change in viscosity with temperature.

Viscosity, kinematic. A comparison (ratio) of absolute viscosity to the specific gravity of a fluid. The standard unit of kinematic viscosity in the English system is the newt (in²/sec); in the cgs system, the stoke (cm²/sec). Kinematic viscosity in stokes, multiplied by specific gravity at the test temperature, equals absolute viscosity in poises.

Type	AVERAGE RELATIVE RATINGS				
	Capacity		Limiting Speed	Permissible Misalignment	
	Radial	Thrust			
SPHERICAL ROLLER BEARINGS					
	SELF-ALIGNING	2.10	0.20 ↔	0.50	± 2°
	SELF-ALIGNING	2.40	0.70 ↔	0.50	± 1°30'
CYLINDRICAL ROLLER BEARINGS					
	SELF-CONTAINED TWO DIR. LOCATING	1.35	locating ↔	1.15	± 0°5'
	SEPARABLE INNER RING ONE DIR. LOCATING	1.55	locating ←	1.15	± 0°5'
	SEPARABLE INNER RING ONE DIR. LOCATING	1.55	locating ←	1.15	± 0°5'
	SEPARABLE INNER RING TWO DIR. LOCATING	1.55	locating ↔	1.15	± 0°5'
	SEPARABLE OUTER RING NON-LOCATING	1.55	0	1.20	± 0°5'
	SEPARABLE INNER RING NON-LOCATING	1.55	0	1.20	± 0°5'
	SELF-CONTAINED NON-LOCATING FULL TYPE	2.10	0	0.20	± 0°5'
	SEPARABLE OUTER RING NON-LOCATING	1.85	0	1.00	0°
	SEPARABLE INNER RING NON-LOCATING	1.10	0	1.00	0°
THRUST BALL BEARINGS					
	SINGLE DIRECTION FLAT RACE	0	0.70 ←	0.10	0° Will accept eccentric misalignment

Type		AVERAGE RELATIVE RATINGS			
		Capacity		Limiting Speed	Permissible Misalignment
		Radial	Thrust		
RADIAL BALL BEARINGS					
	CONRAD TYPE	1.00	0.75 ↔	1.00	±0° 15'
	MAXIMUM TYPE	1.40	0.25 ↔	1.00	±0° 3'
	MAGNETO TYPE	0.80	0.50 ←	1.00	±0° 5'
	ANGULAR CONTACT 20°/40°	20°	1.15 ←	1.10	±0° 2'
		40°	1.00 ↔	1.00	±0° 2'
	AIRCRAFT CONTROL FULL TYPE	high static capacity	0.50 ↔	0.20	0°
	AIRCRAFT CONTROL SELF ALIGNING	0.45	0.05 ↔	0.20	±10°
	MAXIMUM TYPE	1.40	0.20 ↔	1.00	0°
	ANGULAR CONTACT 45°	1.50	1.85 ↔	0.80	0°
	ANGULAR CONTACT 36°	1.85	0.50 ↔ 1.50 ←	0.70	0°
	SELF ALIGNING	0.70	0.20 ↔	1.00	±2° 30'
THRUST BALL BEARINGS					
	SINGLE DIRECTION GROOVED RACE	0	1.50 ←	0.30	0°
	DOUBLE DIRECTION GROOVED RACE	0	1.50 ↔	0.30	0°

Table of Basic Rolling Element Bearing Types—The average relative ratings shown in the four columns provide a comparison of approximate radial and thrust capacity, limiting speed and permissible misalignments.

Appendix D Power Supply Terminology

Accuracy. Used as a specification for the output voltage of power supplies, accuracy refers to the absolute voltage tolerance with respect to the stated nominal output.

Ambient operating temperature (range). The range of environmental temperatures in which a power supply can be safely operated. For units with forced air cooling, the temperature is measured at the air intake.

Bipolar. Having two poles, polarities or directions. Applied to amplifiers or power supplies, it means that the output may vary in either polarity from zero; as a symmetrical program, it need not contain a DC component. (See Unipolar.)

Bridge current. The circulating control current in the comparison bridge. Bridge current equals the reference voltage divided by the reference resistor. Typical values are 1 ma and 10 ma, corresponding to control ratios of 1000 ohms/volt and 100 ohms/volt, respectively.

Calibration programming. Calibration with reference to power supply programming describes the adjustment of the control bridge current to calibrate the programming ratio in ohms per volt. Many programmable supplies incorporate a "calibrate" control as part of the reference resistor which performs this adjustment.

Closed-loop gain (operational gain). The gain, measured with feedback, is the ratio of the voltage appearing across the output terminal pair to the causative voltage required at the input resistor. The closed-loop (operational) gain is denoted by the symbol G in diagrams and equations. If the open-loop gain A is sufficiently large, the closed-loop gain can be satisfactorily approximated by the ratio of the feedback resistor R_f to the input resistor R_1 . (See Open-loop, Loop gain.)

Command reference. In a servo or control system, the voltage or current to which the feedback signal is compared. As an independent variable, the command reference exercises complete control over the system output. (See Operational programming.)

Comparison amplifier. A high gain, noninverting DC amplifier which, in a bridge regulated power supply, has as its input the voltage between the null junction and the common terminal. The output of the comparison amplifier drives the series pass elements.

Comparison bridge. A type of voltage comparison circuit whose configuration and principle of operation resemble a four-arm electrical bridge (Fig. 1). The elements are so arranged that, assuming a balance

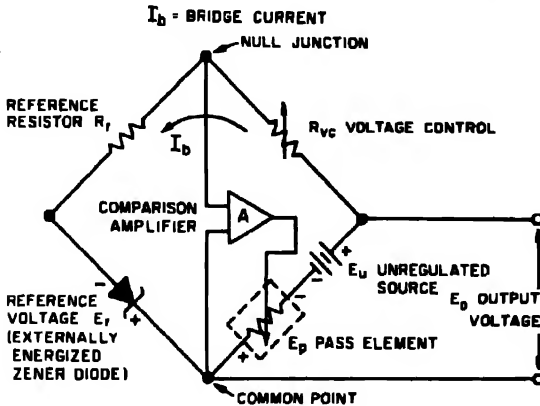


Figure 1. Kepco comparison bridge connected as a voltage regulator.

exists in the circuit, a virtual zero error signal is derived. Any tendency for the output voltage to change in relation to the reference voltage creates a corresponding error signal, which, by means of negative feedback, is used to correct the output in the direction toward restoring bridge balance. (See Error signal.)

Complementary tracking. A system of interconnection of two regulated supplies in which one (the master) is operated to control the other (the slave). The slave supply voltage is made equal (or proportional) to the master supply voltage and of opposite polarity with respect to a common point (Fig. 2).

Compliance extension. A form of master/slave interconnection of two or more current regulated power supplies to increase their compliance voltage range through series connection.

Compliance voltage. The output voltage of a DC power supply operating in constant current mode. The compliance range is the range of voltages needed to sustain a given value of constant current throughout a range of load resistances.

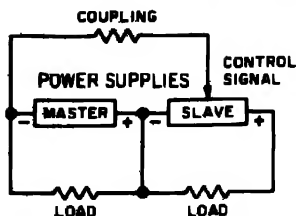


Figure 2. Complementary tracking.

Constant current power supply (current regulator). A power supply capable of maintaining a preset current through a variable load resistance. This is achieved by automatically varying the load voltage in order to maintain the ratio $E_{\text{load}}/R_{\text{load}}$ constant.

Constant voltage power supply (voltage regulator). A power supply that is capable of maintaining a preset voltage across a variable load resistance. This is achieved by automatically varying the output current in order to maintain the product of load current times load resistance constant.

Control ratio. The required change in control resistance to produce a one volt change in the output voltage. The control ratio is expressed in ohms per volt and is reciprocal of the bridge current.

Cooling. In power supplies, the cooling of regulator elements refers to the method used for removing heat generated in the regulating process. Methods include radiation, convection, and conduction or combinations thereof.

Cooling, convection. A method of heat transfer which uses the natural upward motion of air warmed by the heat dissipators.

Cooling, lateral forced air. An efficient method of heat transfer by means of side-to-side circulation which employs blower movement of air through or across the heat dissipators.

Crossover (automatic) voltage current. The characteristic of a power supply that automatically changes the method of regulation from constant voltage to constant current (or vice versa) as dictated by varying load conditions (Fig. 3). The constant voltage and constant current levels can be independently adjusted within the specified voltage and current limits of the power supply. The intersection of constant voltage and constant current lines is called the cross-over point E, I and may be located anywhere within the volt-ampere range of the power supply.

Current cut-off. An overload protective mechanism designed into certain regulated power supplies to automatically *reduce* the load current

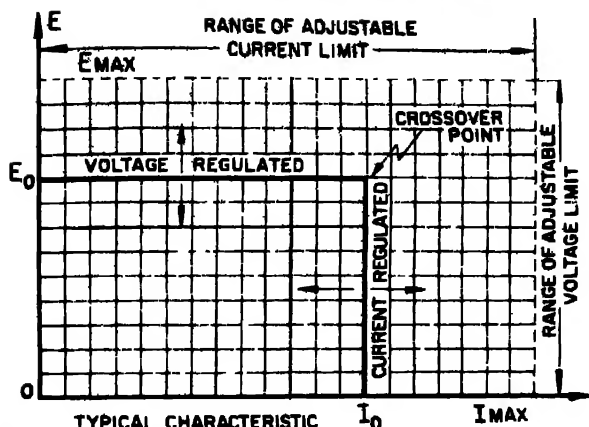


Figure 3. Automatic voltage/current crossover.

as the load resistance is *reduced*. This "negative resistance" characteristic reduces overload dissipation to negligible proportions and protects sensitive loads. See Fig. 4 for the E, I characteristic of a power supply equipped with a current cutoff overload protector.

Current limiting (automatic). An overload protection mechanism which limits the maximum output current to a preset value, and automatically restores the output when the overload is removed. (See Short circuit protection and Fig. 5.)

Current sensing resistor. A resistor placed in series with the load to develop a voltage proportional to load current. A current regulated DC power supply regulates the current in the load by regulating the voltage across the sensing resistor.

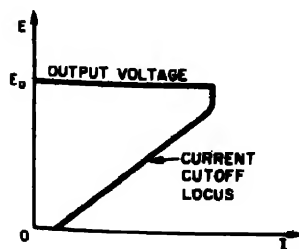


Figure 4. Output characteristic of a power supply equipped with current cutoff overload protector.

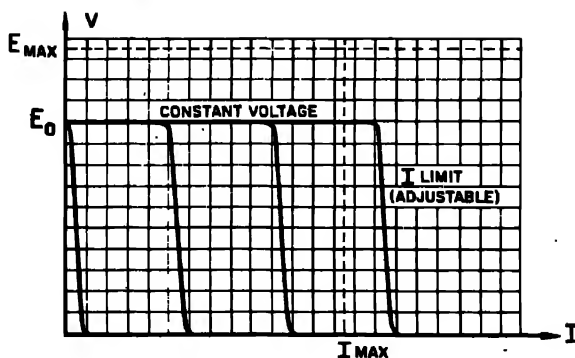


Figure 5. Plot of typical current limiting curves.

Delta minimum. A qualifier, often appended to a percentage specification to describe that specification when the parameter in question is a variable, and particularly when that variable may approach zero. The qualifier is often known as the “minimum delta V ,” or minimum delta I ,” as the case may be.

Drift. See Stability.

Error signal. The error signal is the difference between the output voltage and a fixed reference voltage compared in ratio by the two resistors at the null junction of the comparison bridge; i.e., $\epsilon = E_o - E_r[R_{vc}/R_r]$ (Fig. 1). The error signal is amplified to drive the pass elements and correct the output.

Filters. Filters are RC or LC networks arranged as low pass devices to attenuate the varying component that remains when AC voltage is rectified. In power supplies without subsequent active series regulators, the filters determine the amount of ripple that will remain in the DC output. In supplies with active feedback series regulators, the regulator mainly controls the ripple with output filtering serving chiefly for phase-gain control as a lag element.

Flux-O-Tran. A registered trademark of Kepco, Inc., applied to ferro-resonant voltage regulating transformers of a special design, which are used in many proprietary designs. The Flux-O-Tran, with its resonating capacitor provides a squarewave output (for high rectifier and filter efficiency) whose magnitude is largely independent of the primary voltage amplitude.

Frequency response. The measure of an amplifier or power supply's ability to respond to a sinusoidal program. The frequency response measures

the maximum frequency for full-output voltage excursion. This frequency is a function of the slewing rate and unity gain bandwidth.

Full-wave rectification. In the rectifying process, full-wave rectification inverts the negative half-cycle of the input sinusoid so that the output contains two half-sine pulses for each input cycle. A pair of rectifiers arranged as shown with a centertapped transformer, or a bridge arrangement of four rectifiers and no centertap are both methods of obtaining full-wave rectification. (See Fig. 6.)

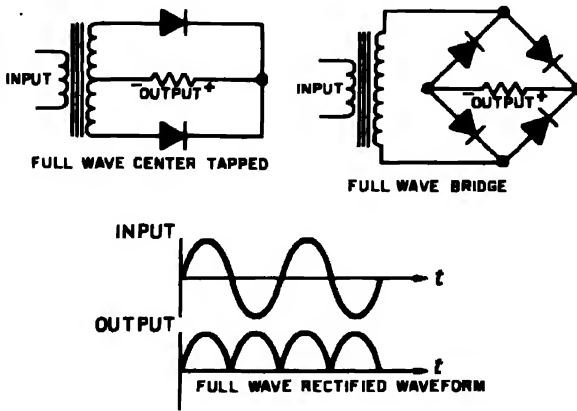


Figure 6. Full-wave rectification.

Half-wave rectification. In the rectifying process, half-wave rectification passes only one-half of each incoming sinusoid, and does not pass the opposite half-cycle. The output contains a single half-sine pulse for each input cycle. A single rectifier, arranged as in Fig. 7, provides half-wave rectification. Because of its poorer efficiency and larger AC component, half-wave rectification is usually employed in noncritical low current circumstances.

High speed regulator. A power supply regulator circuit which, by the elimination of its output capacitor, has been made capable of much higher slewing rates than are normally possible. High speed (HS) regulators are used where rapid step programming is needed, or as current regulators for which they are ideally suited. (See Slewing rate.)

Hybrid. A combination of disparate elements to form a common circuit. In power supplies, the combination of vacuum tubes and transistors in the regulating circuitry.

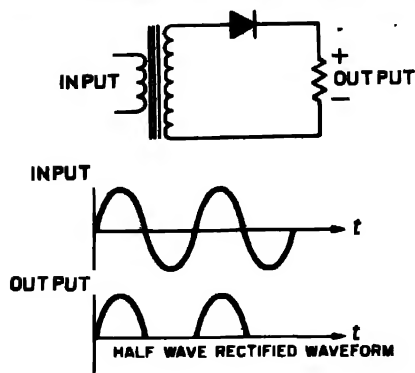


Figure 7. Half-wave rectification.

Inverting amplifier. An amplifier whose output polarity is reversed as compared to its input. Such an amplifier obtains its negative feedback by a connection from output to input, and with high gain is widely used as an operational amplifier. An operational DC power supply can also be described as a high gain inverting amplifier.

Isolation voltage. A rating for a power supply which specifies the amount of *external* voltage that can be connected between any output terminal and ground (the chassis). This rating is important when power supplies are connected in series.

Lag network. Resistance-reactive components, arranged to control phase-gain rolloff versus frequency. Used to assure the dynamic stability of a power supply's comparison amplifier. The main effect of a lag network is a reduction of gain at relatively low frequencies so that the slope of the remaining rolloff can be relatively more gentle.

Lead network. Resistive-reactive components arranged to control phase-gain rolloff versus frequency. Used to assure the dynamic stability of a power supply's comparison amplifier. The main effect of a lead network is to introduce a phase lead at the higher frequencies, near the unity gain frequency.

Linearity, programming. The linearity of a programming function refers to the correspondence between incremental changes in the input signal (resistance, voltage or current) and the consequent incremental changes in power supply output. Direct programming functions are inherently linear for the Kepco bridge regulator, and are accurate to within a percentage equal to the supply's regulating ability.

Line regulation. The maximum steady-state amount that the output voltage or current will change as the result of a specified change in line voltage (usually for a step change between 105-125 or 210-250 volts, unless otherwise specified). Regulation is given either as a percentage of the output voltage or current, and/or as an absolute change, ΔE or ΔI .

Load regulation. The maximum steady state amount that the output voltage or current will change as the result of a specified change in output load, generally from no-load to full-load unless otherwise specified. Regulation is given either as a percentage of the output voltage or current and/or as an absolute change, ΔV or ΔI .

Loop (leakage) current. A DC current flowing in the feedback loop (voltage control) independent of the control current generated by the reference zener diode source and reference resistor. The loop (leakage) current remains when the reference current is made zero. It may be compensated for, or nulled in special applications to achieve a very high impedance (zero current) at the feedback (voltage control) terminals.

Loop gain. A measure of the feedback in a closed-loop system, being equal to the ratio of the open-loop to the closed-loop gains, in db. The magnitude of the loop gain determines the error attenuation and, therefore, the performance of an amplifier used as a voltage regulator. (See Open-loop and Closed-loop gain.)

Master/slave operation. A system of interconnection of two regulated power supplies in which one (the master) operates to control the other (the slave). Specialized forms of the master/slave configuration are used in
 a) *Complementary tracking* (plus and minus tracking around a common point),
 b) *Parallel operation* to obtain increased current output for voltage regulation,
 c) *Compliance extension* to obtain increased voltage output for current regulation.

MTBF Mean time between (or before) failure. A measure of reliability giving either the time before first failure or, for repairable equipment, the average time between repairs. MTBF may be approximated or predicted by summing the reciprocal failure rates of individual components in an assembly.

Null junction. That point on the Kepco bridge at which the reference resistor, the voltage control resistance and one side of the comparison amplifier coincide. The null junction is maintained at almost zero potential and is a *virtual ground*. (See Summing point.)

Offset voltage. A DC potential remaining across the *comparison amplifier's* input terminals (from the null junction to the common terminal) when the output voltage is zero. The polarity of the offset voltage is such as to allow the output to pass through zero and the polarity to be reversed. It is often

deliberately introduced into the design of power supplies to reach and even pass zero output volts.

Open-loop gain. The gain, measured without feedback, is the ratio of the voltage appearing across the output terminal pair to the causative voltage required at the (input) null junction. The open-loop gain is denoted by the symbol A in diagrams and equations. (See Closed-loop and Loop gain.)

Operational power supply. A power supply whose control amplifier has been optimized for signal processing applications rather than the supply of steady-state power to a load. A self-contained combination of operational amplifier, power amplifier and power supplies for higher level operation applications.

Operational programming. The process of controlling the output voltage of a regulated power supply by means of signals (which may be voltage, current, resistance or conductance) which are *operated on* by the power supply in a predetermined fashion. Operations may include algebraic manipulations, multiplication, summing, integration, scaling and differentiation. (See Fig. 8.)

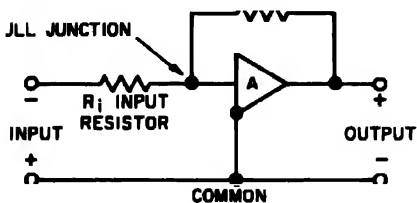


Figure 8. Operational programming.

Output impedance. The effective dynamic output impedance of a power supply is derived from the ratio of the measured peak-to-peak change in output voltage to a measured peak-to-peak change in alternating load current. Output impedance is usually specified throughout the frequency range DC to 100 kc.

Overshoot. A transient rise beyond regulated output limits, occurring when the AC power input is turned on or off, and for line or load step changes. (See Figs. 9, 11a-b.)

Over-temperature protection. A thermal relay circuit which turns off the power automatically should an over-temperature condition occur.

Parallel operation. Voltage regulators, connected together so that their individual output currents are added and flow in a common load. Several

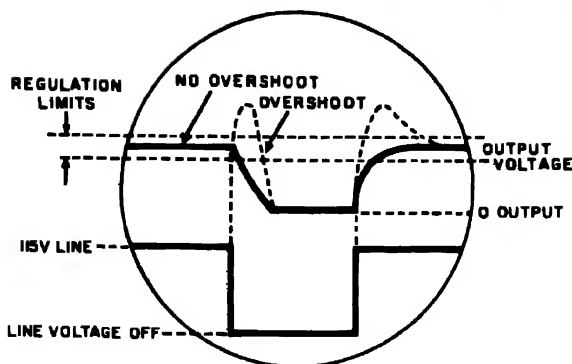


Figure 9. Scope view of turn-off/turn-on effects on a power supply.

methods for parallel connection are used: spoiler resistors, master/slave connection, parallel programming and parallel padding. Current regulators can be paralleled without special precaution.

Parallel padding. A method of parallel operation for two or more power supplies in which their current limiting or automatic crossover output characteristic is employed so that each supply regulates a portion of the total current, each parallel supply adding to the total and "padding" the output only when the load current demand exceeds the capability—or limit setting—of the first supply.

Parallel programming. A method of parallel operation for two or more power supplies in which their feedback terminals (voltage control terminals) are also paralleled. These terminals are often connected to a separate programming source.

Pass element. A controlled variable resistance device, either a vacuum tube or power transistor, in series with the source of DC power. The pass element is driven by the amplified error signal to increase its resistance when the output needs to be lowered or to decrease its resistance when the output must be raised. (See Series regulator.)

Power supply (AC to DC). Generally, a device consisting of a transformer, rectifier and filter for converting AC to a prescribed DC voltage or current.

Programming. The control of any power supply functions, such as output voltage or current, by means of an external or remotely located variable control element. Control elements may be variable resistances, conductances, or variable voltage or current sources. (See Fig. 10.)

Programming speed. Programming speed describes the time required to change the output voltage of a power supply from one value to another. The

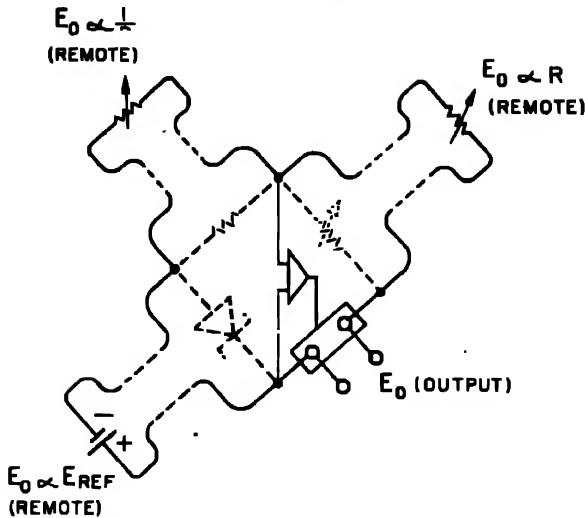


Figure 10. Remote programming connection.

output voltage must change across the load and because the supply's filter capacitor forms an RC network with the load and internal source resistance, programming speed can only be described as a function of load. Programming speed is the same as the "recovery time" specification for *current regulated* operation; it is not related to the recovery time specification for voltage regulated operation.

Recovery time (current regulation). Specifies the time needed for the output current to return to a value within the regulation specification after a step load or line change. For load change, current will recover at a rate governed by the rate-of-change of the compliance voltage across the load. This is governed by the RC time constant of the output filter capacitance, internal source resistance and load resistance. (*See Programming speed.*)

Recovery time (voltage regulation). Specifies the time needed for the output voltage to return to a value within the regulation specification after a step load or line change. Recovery time, rather than response time, is the more meaningful and therefore preferred way of specifying power supply performance, since it relates to the regulation specification. (*See Figs. 11a-b.*)

Regulated power supply. A power supply which maintains a constant output voltage (or current) for changes in the line voltage, output load, ambient temperature or time:

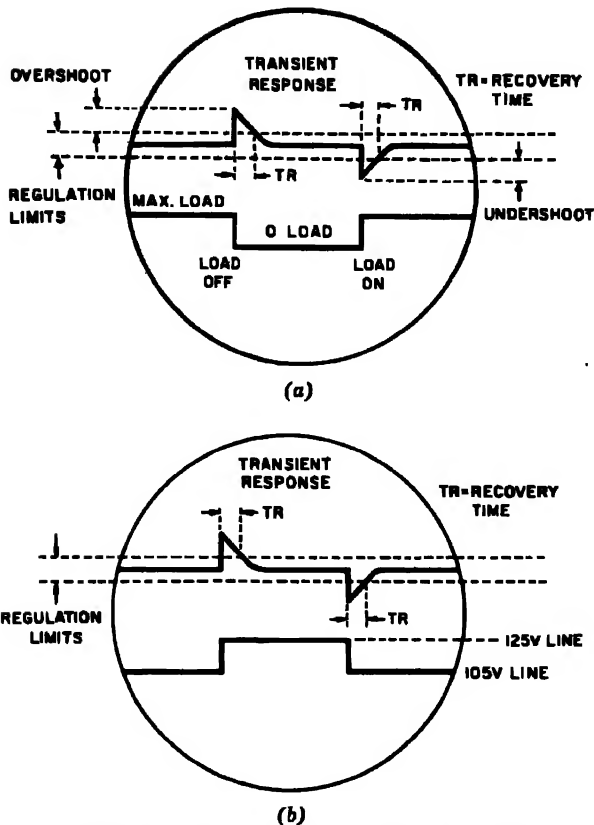


Figure 11. Scope view shows the effects of a step load change (a) and a step line change (b).

Regulation. The maximum amount that the output will change as a result of the specified change in line voltage, output load, temperature or time. Line regulation, load regulation, stability, and temperature coefficient are defined and usually specified separately.

Remote error sensing. A means by which the regulator circuit senses the voltage directly at the load. This connection is used to compensate for voltage drops in the connecting wires.

Response time (time constant). Specifies the time required for a voltage or current excursion to be reduced to 37% of its peak value after a step load or line change. This is not the preferred way of specifying voltage regulator performance. (See Recovery time.)

Resolution. The minimum voltage (or current) increment within which the power supply's output can be set using the panel controls. For continuous controls, the minimum increment is taken to be the voltage (or current) change caused by one degree of shaft rotation.

Ripple. Stated either in peak-to-peak or in *rms* value, ripple specifies the maximum AC component that appears in a DC output. Unless specified separately, ripple includes unclassified noise.

Series operation. The output of two or more power supplies connected together to obtain a total output voltage equal to the sum of their individual voltages. Load current is equal and common through each supply. The extent of series connection is limited by the maximum specified potential rating between any output terminal and ground. (*See Isolation voltage.*) For series connection of current regulators, master/slave (compliance extension) or automatic crossover is used.

Series regulator. A device placed in series with a source of power that is capable of controlling the voltage or current output by automatically varying its series resistance. (*See Pass element.*)

Short circuit protection (automatic). Any automatic current limiting system which enables a power supply to continue operating at a limited current, and without damage, into any output overload including short circuits. The output voltage must be restored to normal when the overload is removed, as distinguished from a fuse or circuit-breaker system which opens at overload and must be closed to restore power. (*See Current limiting, Fig. 5.*)

Shunt regulator. A device placed across the output, which controls the current through a series dropping resistance to maintain a constant voltage or current output.

Slaved tracking. A system of interconnection of two or more regulated supplies in which one (the master) operates to control the others (the slaves). The output voltage of the slave units may be equal or proportional to the output voltage of the master unit. (The slave output voltages track the master output voltage in a constant ratio.) (*See Complementary tracking, Master/slave.*)

Slewing rate. A measure of the programming speed or current-regulator response timing. The slewing rate measures the maximum rate-of-change of voltage across the output terminals of a power supply. Slewing rate is normally expressed in volts per second ($\Delta E/\Delta T$) and can be converted to a sinusoidal frequency-amplitude product by the equation $f(E_{pp}) = \text{slewing rate}/\pi$, where E_{pp} is the peak-to-peak sinusoidal volts. Slewing rate = $\pi f(E_{pp})$. (*See High speed regulator.*)

Spoiler resistors. Resistors used to *spoil* the load regulation of regulated power supplies to permit parallel operation when not otherwise provided for.

Stability, long term (LTS). The change in output voltage or current as a function of time, at constant line voltage, load and ambient temperature (sometimes referred to as drift).

Step line voltage change. An instantaneous change in line voltage (e.g., 105-125 V AC); for measuring line regulation and recovery time.

Step load change. An instantaneous change in load current (e.g., zero to full load) for measuring the load regulation and recovery time.

Summing point. (*See* Null junction.) The null junction is called a summing point because, as the input to a high gain DC amplifier, operational summing can be performed at this point. As a virtual ground, the summing point decouples all inputs so that they add linearly in the output, without other interaction. (*See* Operational programming.)

Temperature coefficient (TC). The percent change in the output voltage or current as a result of a 1°C change in the ambient operating temperature (% per °C).

Temperature, operating. The range of environmental temperatures in which a power supply can be safely operated (typically, -20°C to +50°C). [*See* Ambient operating temperature (Range).]

Temperature, storage. The range of environmental temperatures in which a power supply can be safely stored (typically, -40°C to +85°C).

Unipolar. Having but one pole, polarity or direction. Applied to amplifiers or power supplies, it means that the output can vary in only one polarity from zero and, therefore, must always contain a DC component. (*See* Bipolar.)

Unity gain bandwidth. A measure of the gain-frequency product of an amplifier. Unity gain bandwidth is the frequency at which the *open-loop gain* becomes unity, based on a 6 db per octave crossing. [*See* Fig. 12, Typical Gain-Frequency (Bodé) Plot.]

VIP. A model designation of Kepco, Inc., applied to a group of load protectors: (V) Voltage, (I) Current, (P) Protectors. The VIP devices provide overvoltage, undervoltage and over/under current sensing and protection circuits.

VIX,[®] indicators. Voltage/Current Crossover Indicators. VIX indicators are a pair of small mode lamps on the front panel of automatic crossover power supplies. One lamp lights during voltage regulated operation of the power supply, the other, during current regulation operation.

VIX signal. A keyed voltage, whose polarity is an indication of power supply output voltage/current regulation mode. The polarity abruptly

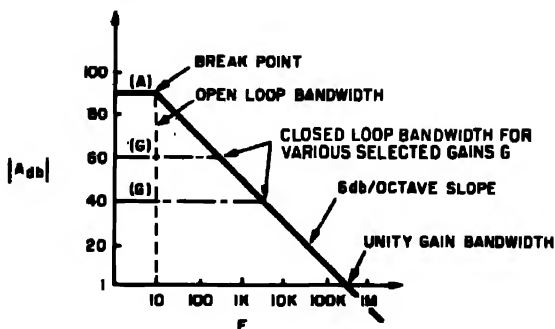


Figure 12. Gain-frequency (Bode) plot.

reverses at the crossover point and can be used to actuate external mechanisms such as lamps, alarms, etc.

Voltage corrector. An active source of regulated power placed in series with an unregulated supply to sense changes in the output voltage (or current); also to correct for the changes by automatically varying its own output in the opposite direction, thereby maintaining the total output voltage (or current) constant. (See Fig. 13.)

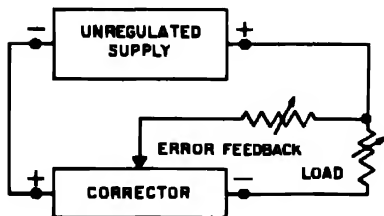


Figure 13. Circuit used to sense output voltage changes.

Voltage reference. A separate, highly regulated voltage source used as a standard to which the output of the power supply is continuously referred.

Warmup time. The time (after power turn-on) required for the output voltage, or current, to reach an equilibrium value within the stability specification.

Appendix E Operational Amplifier Nomenclature

Admittance. For sinusoidal signals, the incremental ratio of a current to a voltage. A *self-admittance* describes the current-voltage relationship in a two-terminal element; a *driving-point* admittance relates the current into a terminal to the voltage between that terminal and common; a *trans-admittance*, in general, relates a current into any terminal of a circuit to a voltage between any pair of terminals.

Arbitrary function fitter. A circuit having an output voltage or current that is a presettable, adjustable, (usually non-linear) function of the input voltage(s) or current(s) fed to it.

Bandwidth. Generally, the frequency range over which a particular transfer characteristic (i.e., gain, attenuation, phase-shift, etc.) is maintained between two sets of terminals (i.e., input and output). In an operational amplifier, the frequency range over which the open-loop gain exceeds unity. In an operational-amplifier *circuit*, the frequency range over which the (small-signal) loop gain maintains the desired response, to the desired accuracy.

Bias circuit. A (fixed or adjustable) circuit that is used to set amplifier (zero-signal) input-current or input-voltage level to an arbitrary value (normally zero). May be "temperature-compensating". . . i.e., able to track the variation of amplifier input voltage or current with temperature, more or less perfectly.

Booster amplifier. A circuit used to increase the output current or voltage capabilities of an operational amplifier circuit, without loss of accuracy (ideally) or inversion of polarity. Usually applied *inside* the loop, for accuracy.

Bound circuit. A circuit designed to limit the excursion of a signal. The limit value it establishes may be nominal (when used for protection), or highly-precise (when used operationally).

Chopper. A circuit or device for interrupting (or at least modifying) a low-frequency signal path at a constant rate (i.e., carrier frequency), producing a wave, which is modulated by the DC signal magnitude, preserving the polarity of the DC signal. Generally associated with a *synchronous demodulator*, following amplification.

Common-mode error (CME) (referred to the input). A (generally) small offset voltage appearing between the input terminals of a differential operational amplifier, as a function of the *common-mode* voltage.

Common-mode rejection ratio (CMRR). The ratio of common-mode voltage to common-mode error in a differential amplifier circuit.

Common-mode voltage. The voltage between the output signal return and (in this book) the positive input terminal of a differential operational amplifier.

Comparator, precision. A high-gain amplifier circuit whose output changes decisively between two definite levels whenever the sum of the input voltages changes sign.

Controller. A portion of a feedback system in which the unregulated unbalance (or "error") signal is operated on by adjustable dynamic elements (proportional, integral, derivative, lead-lag, etc.) to affect the manipulated variable in such a way that desirable response criteria for the loop (stability, speed, accuracy) may be met.

Current bias. *See* Bias.

Current pump. A circuit that drives, through an external (load) circuit, an adjustable variable or constant value of current, regardless of the reaction of that load to the current, within rated limits of current, voltage, and load impedance.

Damping circuit. A circuit used to limit, control, or prevent dynamic instability (oscillation or "ringing") in a closed-loop active circuit, or in a complex passive network having appreciable second-order (or higher-order) response.

Dead zone. A range in which no output change is produced by substantial input variations; a circuit element having such response (or lack of response).

Differential-input amplifier. One in which the output is (ideally) a function only of the *difference* between the signals applied to its two inputs, both signals being measured with respect to a common "low," or "ground" reference point.

Differentiator. Ideally, a circuit having a response (output) proportional to the time-derivative(s) of one or more input signals.

Diode bounding. A form of *Bounding* in which the nonlinear conducting

properties of a diode (or diodes) are used to accomplish the magnitude-limiting action.

DC beta. The DC current gain of a transistor; the ratio of the collector current to the base current that caused it, measured at constant collector to emitter voltage.

Drift. A gradually-developing deviation in any voltage, current or impedance. For an operational amplifier, a gradually-developing change in the offset voltage or in either or both of the offset currents. Also, the bottom-most frequency range of the noise spectrum.

Electrometer amplifier. An amplifier circuit having sufficiently low-current drift and other noise components, sufficiently low amplifier input-current offsets, and adequate power and current sensitivities to be usable for measuring current variations considerably less than 10^{-12} A.

Emitter-follower. In principle, a single-transistor amplifier in which the load is connected between the emitter and signal ground, so that the base-to-emitter-to-ground path (for the input signal) contains 100% negative feedback of the output voltage. The collector is returned (in principle) directly to the power supply. The gain is very nearly unity and the output signal is *not* inverted (i.e., it "follows" the input).

Error-factor, finite-gain. That factor by which the "ideal" closed-loop response expression must be multiplied to yield the response for an amplifier with *finite* gain, A , rather than infinite gain, as is assumed in computing the "ideal" response.

Fault current. The current that may flow in any part of a circuit or amplifier under (specified) abnormal conditions.

Follower-with-gain. A follower (*which see*) in which only a *part* of the output voltage is fed back in series opposition to the input signal . . . hence, closed-loop gain greater than unity is obtained over the rated range of operation.

Feedback circuit. A causal circuit configuration in which (for the simplest circuit) the input and the output variables are combined and together determine the output.

Flicker. Noise in an amplifier, of higher frequency than drift, but lower than power-line or chopper-drive frequency noise. Also called "jitter" or "wobble."

Follower. A circuit in which the output of a high-gain amplifier is fed directly back to its negative input. The input signal is reproduced without polarity reversal. *See also* Emitter follower and Follower with gain.

Flip-flop. *See* Multivibrator, bistable.

Frequency, angular. The rate of change of the angle of a sine wave, expressed in radians per second, where 2π radians (360°) = 1 alternation (cycle).

Frequency, break. In a plot of log gain (attenuation) vs. log frequency, the frequency at which the asymptotes of two adjacent linear slope segments meet.

Gain-bandwidth product. (1) The product of a specific frequency and the gain of a circuit, amplifier, or system at that frequency. (2) For an operational amplifier, or any other circuit or device having the special property that its gain is inversely proportional to frequency, the G. B. P. is equal to the frequency at which the gain falls (by extrapolation) to unity.

Gain, closed-loop. The response of a feedback circuit to a voltage inserted in series with the *amplifier* input. Also, the "noise gain."

Gain, loop. In an operational amplifier circuit, the product of the transfer characteristics of all of the elements (active or passive) encountered in a complete trip around the loop, starting at any point and returning to that point.

Gain, open loop. The ratio of the (loaded) output of an Amplifier to its net input, at any frequency. Usually implies *voltage gain*.

Gate, precision. A circuit that may be switched from closed- to open-circuit or vice versa without error (time, bias, impedance) in response to a command signal (voltage or current).

Ground chassis. The potential (assumed uniform) of the (metallic) structure on or in which the circuit is built.

Ground, power-common. The potential of the terminal or circuit point to which the output of a power supply (and often an amplifier output load) is returned (i.e., power-supply "zero").

Ground, signal (or high-quality). The potential of a terminal or circuit point to which all signal voltages are referenced, by convention or arbitrary assumption. Usually the signal-return of the lowest-level signal in a system.

"Hold" mode. In integrators or other charge-storage circuits, a condition (or time-interval) in which input(s) are removed and the circuit is commanded (or expected) to maintain constant output.

Hysteresis. A form of non-linearity in which the response of a circuit to a particular set of input conditions depends, not only on the instantaneous values of those conditions, but also on the immediate past (recent history) of the input and output signals. Hysteretical behavior is characterized by inability to "retrace" exactly on the reverse swing a particular locus of input/output conditions.

Idling current. The zero-signal power supply current drawn by a circuit, or by a complete amplifier. Also called "Quiescent" current.

Impedance, input, common-mode. The (internal) impedance between either one of the input terminals of a differential operational amplifier and signal ground.

Impedance, input, differential. The (internal) impedance observed between the input terminals of an operational amplifier.

Impedance, negative. In general, the driving-point impedance of circuit in which a current increase produces a voltage decrease (and vice versa); for a *negative admittance*, a voltage increase produces a current decrease, and vice versa.

Integrator. A circuit having a response (output) proportional to the time-integral of one or more input signals.

Inverter, voltage. A circuit having a response (output) proportional to a constant (the *gain*) times the input signal, but opposite in sign to it. In a unity-gain inverter, the output is (-1) times the input.

Lag (noun). A delayed-response characteristic, or a circuit having such a delayed response. Usually 1st-order lag is implied unless otherwise specified. (verb). To respond to a stimulus in delayed fashion.

Lag-lead (lead-lag). A circuit whose response includes lag components and their derivatives.

Leakage. (Unwanted) current flow through a nominally-blocked (non-conducting) circuit or circuit element due to imperfections in its blocking behavior.

Limits. See Bounds.

Memory, peak or valley readout. A circuit in which the output remains at the condition corresponding to the most positive (least negative) or vice versa input signal since the circuit was set to initial conditions, until reset to those conditions.

Multiplier, quarter-square. A circuit that achieves true four-quadrant multiplication by taking advantage of the mathematical relationship that the product of two variables is equal to one quarter of the difference of the squares of the sum and the difference of the variables.

Multivibrator, astable (free-running). A circuit having two momentarily stable states, between which it continuously alternates, remaining in each for a period controlled by the circuit parameters, and switching rapidly from one to the other.

Multivibrator, bistable (flip-flop). A circuit having two stable states, in either one of which it will stay indefinitely, until triggered appropriately, immediately after which it switches to the other state.

Multivibrator, monostable⁻ (one-shot). A circuit having only one stable state, from which it can be triggered to change state, but only for a predetermined interval, after which it returns to the original state.

Noise, amplifier. All spurious or unwanted signals, random or otherwise, that can be observed in a completely isolated amplifier in the absence of a genuine input signal. (*See also* Drift and Flicker.)

Null detector. A comparator (*which see*) having zero reference voltage. A *graded-null* detector has decreasing sensitivity away from the null.

Offset current. A DC error current appearing at either input terminal of a DC amplifier.

Offset voltage. A DC error voltage appearing in series with either input terminal of a DC amplifier.

Offset, end-resistance. In potentiometers, the residual resistance between a terminal and the moving contact, at a position corresponding to full rotation against that terminal.

Passive network. A network whose net influx (or efflux) of available energy is stored or dissipated within the network. There may be no sources of energy other than those explicitly bookkept as influxes.

Phase characteristic. A graph of phase shift vs. frequency, assuming sinusoidal input and output.

Phase shift. Phase angle between two related variables in a circuit (usually input and output voltage) when excited by sinusoidal signal(s).

Rate limiting. Non-linear behavior in an amplifier due to its limited ability to produce large, rapid changes in output voltage (slewing)—restricting it to rates of change of voltage lower than might be predicted by observing the small-signal frequency response.

Reset mode (set mode). In integrators, memories, or other charge-storage circuits, a state (or time-interval) in which the circuit is forced to return to a set of initial conditions, removing all record of its previous condition.

RMS value. The square root of the time average of the square of a variable signal.

Roll-off. The decrease in magnitude of gain with frequency. Typical roll-off (low-pass) of a circuit for which the dominant lag is first-order is 6 db per octave (inversely proportional to frequency). “Steep” roll-off might be at 12 or 18 db per octave (proportional to the inverse square or cube of frequency) or more.

Saturation voltage. Generally, the voltage excursion at which a circuit self-limits . . . i.e., is unable to respond to excitation in a proportional

manner. In operational amplifiers, the output-voltage saturation limits may be imposed by any stage, from the input to the output, depending in part on the external loading and feedback parameters.

Scaling. Adjusting the coefficient of a circuit to each of its one or more input signal terminals. The *relative* scaling (of one input to another) is called "weighting." In computing, relating problem variables to machine variables.

Slewing rate. *See* Rate limiting.

Soakage. The disability of a capacitor to come up to voltage instantaneously, without voltage lag or creep, during or after charging. The lower the soakage, the lower the lag and creep.

Stabilizer. (DC) A circuit that uses a chopper and preamplifier to maintain the net offset near zero at the input terminals of an operational amplifier. *See* Chopper-stabilized. (Dynamic) An element or circuit employed to promote dynamic stability, also *damp*er.

Subtractor. An operational amplifier circuit in which the output is proportional to the difference between its two input voltages (or between the net sums of its positive and negative inputs).

Time integral. The definite integral of a variable over an interval of time. Also the area under a curve of a function of time during that period. Divide it by the time interval to obtain the average value of the argument over that period.

Track-hold memory. A circuit that, in its "track" mode, develops an output that follows (ideally) the input exactly, or is proportional to it; and then, in its "hold" mode, maintains the output constant (ideally) at the value it had at the instant the circuit was commanded to change from "track" to "hold."

Transconductor. A device that produces a current at a given point in the circuit (usually an amplifier's summing point) as a function of a voltage or voltages, usually at its input or output.

Transdiode. A transistor so connected that base and collector are actively maintained at equal potentials, though not connected together. The logarithmic transfer relationship between collector current and base-emitter voltage very closely approximates that of an *ideal* diode.

Uncertainty, input. In an Operational amplifier, the algebraic sum of all the factors, including environmental and time effects, that contribute to the non-ideal behavior of the input circuit.

Weighting. *See* Scaling.

Index

- Acceleration, transducers, 103–119
 - bubble type, 110
 - density, 105–106
 - force-balance, 111
 - navigational, 108
 - pendulous floated, 114
 - pendulous integrating, 111–114
 - piezoelectric, 109, 115–118
 - piezoresistive, 109, 118–119
 - random, 106
 - selection, 107–119
 - vertical reference, 106–110
 - vibrating beam, 114–115
- Altimeter, 123–124
- Amplifiers, 2, 78
 - operational, 429–448; *see also* Operational amplifiers
- Analogue-to-digital converters, 7
 - capacitor charging, 8
 - counter, ramp, 12
 - discrete voltage, 12
 - emitter-follower, 15
 - operational amplifier, 15
 - parallel, 18
 - pulse width modulator, 8
 - serial/parallel, 20
 - successive approximation, 13
 - up-down integrator, 10
 - voltage to frequency, 8
- Analogue versus digital sensors, 2
- Aperture rate, 6
- Applanation transducer, 122–123
- Batteries, 345–346, 347, 348
- Bearings, 346–350, 395, 400–417
 - barrier films, 405–406
 - cryogenic, 406–409
 - dry film lubrication, 402–404
 - hollow balls, 416–417
 - integral gear race, 413–415
 - low noise, 413–414
 - Roundway, 416–417
 - selection, 349–350
 - shaking force, 349, 413
 - space environment, 400–402
- Beryllium, 411–412
- Beta gage, 71
- Biot Modulus, 125
- Brakes, 391–395
 - brake-clutch combinations, 396–400
 - magnetic, 393
 - mechanical, 392
 - rating, 393–395
 - selection, 391–395
- Capacitive transducers, 57–59
- Clocks, 197–200
- Clutches, 374–390
 - axial, 376–378
 - centrifugal, 380
 - cone, 378
 - disk, 383
 - eddy current, 386–387
 - electric, 381–390
 - friction type, 376
 - hysteresis, 383–385
 - magnetic particle, 385–387
 - multidisk, 378, 383
 - multiple tooth, 376
 - one-way, 379

- positive type, 374–376
 - rim, 376–378
 - spiral, 375
 - sprag, 379–380
 - spring, 380–381
 - square jaw, 375
 - summary, 388–389
 - tooth, 383
 - torque limiting, 380–381
 - torque-slip curve, 382
- Dashpot, 177–180**
- Differentiators, 209–211**
- electronic, 210
 - rate of climb indicator, 211
- Digital-to-analogue, converter, 20**
- design parameters, 28
 - ladder, 23
 - null summing, 26
 - single valued current, 25
 - weighted current, 25
 - weighted resistor, 23
- Displays, output, 4**
- Electromagnetic interference, 356**
- Encoders, 136–174**
- absolute versus incremental, 142
 - ambiguity, 146–148
 - applications, 169–174
 - brush, 137–139, 155–156
 - coding, 143–144
 - Gray code, 144, 147
 - least significant digit, 146
 - light sources, 151–152
 - linear, 167–169
 - magnetic, 140–142, 156–159
 - most significant digit, 146
 - optical, 139–140, 157–166
 - readout devices, 152
 - sine-cosine, 166–167
 - U-scan, 148
 - V-scan, 147–148
- Environmental factors, 451–455**
- acceleration, 454
 - altitude, 454
 - contamination, 455
 - corrosion, 455
 - shock loading, 453
 - vibration, 454
- Escapements, 197–200**
- balance wheel, 197–199
 - run away, 200
- Fiber optics, 63–64**
- Gearing, 341–342**
- materials, 341–342
- Gyro, 202–205**
- Integrators, 190–209**
- area, 193
 - ball and disk, 193–195
 - clocks, 197–200
 - digital voltmeter, 207
 - drag-cup, 202
 - electromechanical, 200–205
 - electronic, 205–207
 - mechanical, 191–200
 - mechanical filter, 202
 - motor-counter, 202
 - motor-tachometer, 202
 - pneumatic, 207–208
 - pulse count, 192
 - pulse rate, 201–202
 - pulse width, 201
 - rate gyro, 202–205
 - selection, 208–209
 - servo, 202–203
 - velocity switch, 196
- Interferometers, 66–70**
- laser, 66
 - moiré, 67, 68
 - microwave, 69–70
- Linear displacement transducer, 32**
- Linear variable differential transformer, 49–54**
- linearity, 54
 - load impedance, 53
 - null voltage, 54
 - phase characteristics, 53
 - resolution, 53
 - sensitivity, 52
 - temperature characteristics, 55
- Lubrication, 350–356**
- barrier film, 405–406
 - cryogenic bearings, 406–409
 - dry film, 354–356, 402–404
 - flash point, 353
 - grease selection, 353–354

- hydrodynamic, 350
- oil selection, 351–353
- pour point, 352
- viscosity, 351–353
- viscosity index, 352
- Motors, 323–360**
 - alternating current type, 332–338
 - capacitor start, 335–336
 - hysteresis, 336
 - induction, 332
 - inertial damped, 337
 - linear type, 337–339, 340
 - polyphase, 332–333
 - repulse, 336
 - servo type, 336–337
 - shaded pole, 336
 - single phase, 333–337
 - split phase, 333–335
 - universal, 336
 - viscous damped, 337
 - battery power, 345–346, 347, 348
 - bearings, 346–350
 - direct current type, 323–332
 - brushless torque type, 328–329
 - ceramic PM shunt, 326–327
 - compound type, 331
 - PM shunt, 325–326
 - printed circuit type, 329–330
 - series, 330–331
 - shunt, 323–325
 - torque type, 327–328
 - duty cycle, 360
 - dynamic braking, 359
 - electromagnetic interference, 356
 - gearing, 341–342
 - lubrication, 350–356
 - noise, 358
 - overload protection, 359
 - reversing, 358–359
 - speed regulation, 344–345
 - thermal considerations, 342–344
- Operational amplifiers, 429–448**
 - chopper circuit, 436–438
 - circuitry, 430–432
 - closed loop gain, 432, 445
 - common mode error, 440
 - common mode rejection ratio, 441–442
 - current offset, 434
 - discrete, 446
 - frequency response, 444
 - integrator, 433
 - hybrid, 447
 - monolithic, 447
 - noise bandwidth, 438
 - noise factor, 439–441
 - open gain, 443
 - settling time, 446
 - slewing, 446
 - summing, 433
 - voltage drift, 434–438
 - voltage offset, 434
- Optical transducers, 60–70**
 - amplifiers, 63
 - detectors, 62
 - fiber optics, 63–64
 - interferometers, 66–70
 - light sources, 60–61
- Outgassing, 452**
- Photoconductive cells, 62**
- Photovoltaic cells, 62**
- Potentiometers, 299–322**
 - AC usage, 38
 - carbon film, 37
 - cermet, 35
 - classification, 300–307
 - construction, 307–314
 - electrical characteristics, 314–318
 - electrical noise, 37
 - failure mode, 38
 - ganged, 301
 - hot-molded carbon, 37
 - multiturn, 301
 - nonlinear, 301–303
 - power rating, 37
 - resolution, 34
 - sealing, 37
 - selection, 37, 318–321
 - setting, 37
 - single turn, 301
 - temperature coefficient, 37
 - thin metal film, 37
 - trimmer, 301
 - wire wound, 35
- Power supplies, 417–429**
 - bridge type, 424–425
 - categories, 418
 - constant current, 425–427

- SCR type, 423–424
 - selection, 427–429
 - series regulator, 419–421
 - shunt regulator, 421–423
- Pressure switch, 124
- Pressure transducers, 119–124
 - applanation, 122–123
 - bellows, 120
 - bourdon tube, 119
 - diaphragm, 121
- Proximity inductive transducer, 55, 56
- Pyrometer, 132–134
- Quantizing, 5
- Radiometer, 133
- Relays, 215–243
 - actuation and release times, 233–235
 - AC versus DC, 214
 - classification, 213–214
 - construction, 222–228
 - contact protection, 226
 - definitions, 217–222
 - electromechanical versus solid state, 216, 217
 - loading, 235
 - optoelectronic type, 227–228, 242–243
 - polarized, 215
 - race, 233
 - reed, 227, 236–240
 - resonant reed, 215–216
 - RFI suppression, 235–236
- Resolvers, 81–90
 - applications, 85–90
 - computing, 83–84
 - synchro, 84
 - theory, 82
- Servo, amplifier, 78
- Servosystems, analogue, 3
 - digital, 3
- Slip rings, 276–299
 - bearings, 294
 - brushes, 281–285
 - design, 278–293
 - drum, 278–280
 - environmental considerations, 298–299
 - instrument type, 295–296
 - lubrication, 292–294
 - noise, 295
 - pancake, 278–280
 - power type, 296–297
 - RF and VF type, 297–298
- Solenoids, 361–374
 - AC versus DC types, 370
 - auxiliary circuits, 371–373
 - configuration, 361
 - construction, 362–368
 - dual-coil type, 371
 - duty cycle, 368
 - holding circuits, 371
 - index number, 367
 - pulsing networks, 373
 - push-pull type, 370
 - rotary, 373–374
 - selection, 368–371
- Space environment, 400–404
 - adhesion, 403
 - cohesion, 403
 - metallurgy, 409–412
- Stepper motors, 255–272
 - applications, 265–270
 - definitions, 270–271
 - drive circuit, 264
 - PM versus variable reluctance, 263–264
 - types, 256–262
- Stepper switches, 244–255
 - applications, 252–256
 - circuitry, 245–249
 - construction, 244–245
- Strain, gages, 38–48
 - adhesives, 45
 - bonded, 38, 42
 - carrier material, 43
 - circuitry, 48
 - current capacity, 41
 - filament, 42
 - gage factor, 42
 - leads, 43
 - range, 40
 - selection, 42
 - semiconductor type, 46, 47, 48
 - temperature compensation, 43
 - unbonded, 38
 - wire and foil types, 42
- Synchros, 72–81
 - configuration, 81
 - control type, 75–78
 - differential transmitter, 76
 - transformer, 76

- transmitter, 76
- transolver, 78
- electrical zero, 80
- envelope, 81
- friction, 81
- input impedance, 79
- input power, 79
- moment of inertia, 81
- motor, 72, 76
- null voltage, 80
- output impedance, 80
- principle of operation, 74
- receiver, 72, 76
- run-out, 81
- selection, 79
- sensitivity, 79
- torque type, 75–76
- transformation ratio, 79
- transmitter, 72
- Systems error analysis, 458–466
 - scheduling, 467
 - tests, 466–467
- Tachometer (AC type), 94–99
 - applications, 97–99
 - damping, 95
 - fundamental voltage, 95
 - generator, 94
 - linearity, 96
 - noise, 95
 - precision, 95
 - rate generator, 95
 - scale factor, 95
 - temperature control, 96
- Tachometer (DC type), 100
- Temperature transducers, 124–134
 - pyrometer, 132–134
 - resistance type, 128
 - surface temperature gradient, 126
 - thermistors, 129–132
 - thermocouples, 127–128
 - time constant, 126
- Thermocouples, 127–128
- Thermistors, 129–132
- Timers, 176–212
 - electromechanical, 183–186
 - electronic, 187–190
 - electrothermal, 180–183
 - magnetic, 187
 - pneumatic, 177–180
- Transducers, 34–134
 - acceleration, 103–119
 - angular, 72–100
 - beta gage, 71
 - capacitive, 57–59
 - inductive, 49
 - interferometers, 66–70
 - optical, 60–65
 - pressure, 119–124
 - proximity inductive, 55, 56
 - resistive, 34
 - resolvers, 81–90
 - strain gages, 38–40
 - synchros, 72–80
 - temperature, 124–134
 - variable reluctance, 56
 - velocity, 90–103
- Variable reluctance transducers, 56
- Velocity switch, 196
- Velocity transducers, 90–103
 - pick-up, 92–93, 102–103
 - tachometer, 94–100
- Vibration, random, 105–106
- Weibull plot, 410–411
- Wheatstone bridge, 48–49
- White noise, 106

